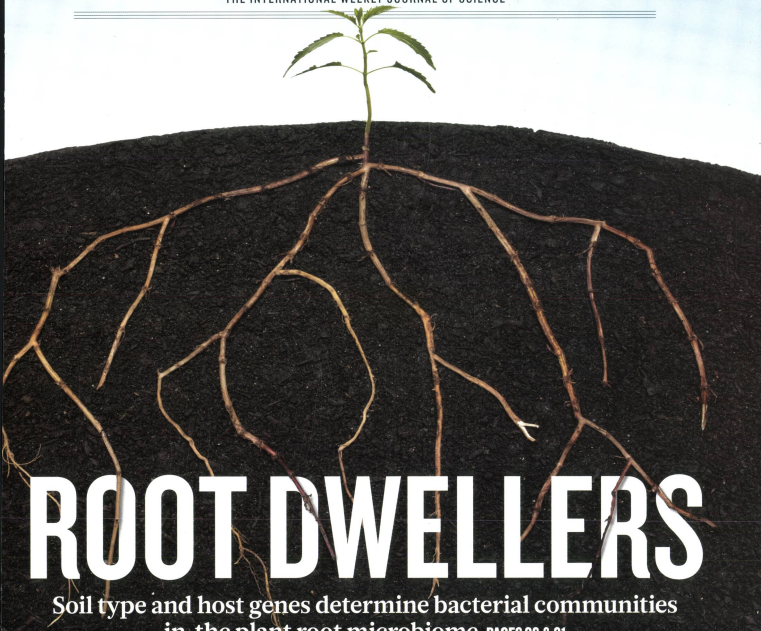
nature

THE INTERNATIONAL WEEKLY JOURNAL OF SCIENCE



ROOT DWELLERS

Soil type and host genes determine bacterial communities in the plant root microbiome PAGES 86 & 91

SOCIAL SCIENCE

HISTORY MAKERS

Mining the past to predict future conflicts

PAGE 24

SPACE RESEARCH

UNIVERSITIES IN SPACE

NASA should spend more on academia's talents

PAGE 27

CAREERS

TEACHING INTEGRITY

Research mentors must set high standards

PAGE 121

NATURE.COM/NATURE

2 August 2012 £10

Vol. 488, No. 7409

9 770028 083095 31>

SCIENTIFIC REPORTS

Report Card

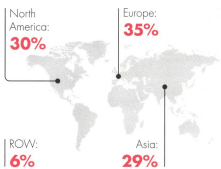
In the first year of publication*

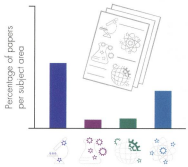
SPEED

99 days on average from submission to publication

DIVERSITY

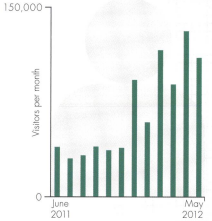
457 papers published in the first year





VISIBILITY

848,000 total visitors



1.89 million total page views

3092 average page views per paper

COMMUNITY

Facebook likes: **32,000+**

93% of authors would submit again†

Keep your research moving. Submit to *Scientific Reports*!

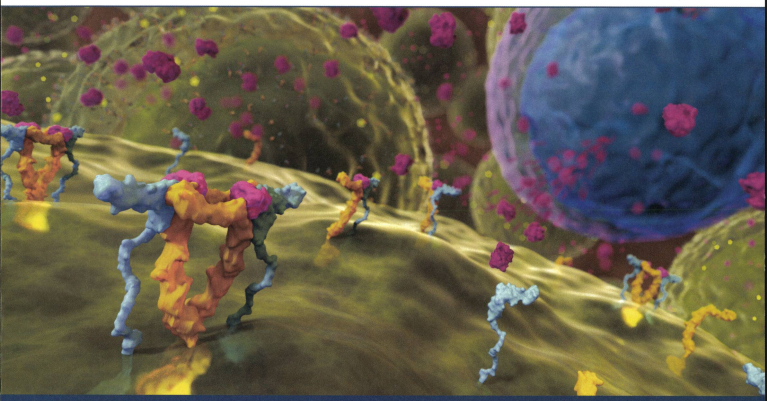
RELIABLY
ACCURATE CELL COUNTS
IN 30 SECONDS.
(BETTER WORKFLOW FOR A BETTER WORKDAY.)

TC20™ AUTOMATED
CELL COUNTER

The TC20™ automated cell counter with built-in auto-focus gives you highly accurate and reproducible mammalian cell counts in a fraction of the usual time—without the need to dilute your sample. So you can avoid the tedium of manually counting cells and move on to other important tasks. It all adds up to a better workflow for your workday.

Watch the video demo at bio-rad.com/ad/TC20

BIO-RAD



The Highest Quality
Cytokines & Growth Factors

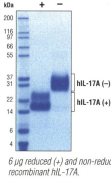
...from Cell Signaling Technology

The world's highest quality antibody provider has now extended its expertise to cytokine production.

- Produced and bioassayed in house with the highest purity and bioactivity.
- Comparison of multiple lots, stringent product specifications, and rigorous quality control ensures maximum lot-to-lot consistency.
- Endotoxin levels are routinely tested and are less than 0.01 ng/µg cytokine.
- Many products are produced in mammalian cells to maximize natural conformation and glycosylation.
- Multi-milligram quantities available.

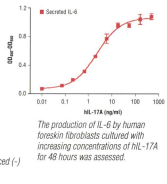
Human Interleukin-17A (hIL-17A) #8928

Purity



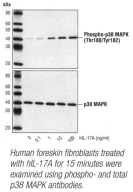
6 µg reduced (+) and non-reduced (−) recombinant hIL-17A.

Bioactivity



The production of IL-6 by human foreskin fibroblasts cultured with increasing concentrations of hIL-17A for 48 hours was assessed.

Downstream Signaling



Human foreskin fibroblasts treated with hIL-17A for 15 minutes were examined using phospho- and total p38 MAPK antibodies.

for quality products you can trust...

www.cellsignal.com

Cell Signaling TECHNOLOGY®

For Research Use Only. Not For Use In Diagnostic Procedures.

Orders (toll-free) 1-877-616-2355 | **Inquiries** info@cellsignal.com | **Environmental Commitment** eco.cellsignal.com

© 2007 Cell Signaling Technology, Inc. Cell Signaling Technology® is a trademark of Cell Signaling Technology, Inc.

CONTENTS

2 August 2012 / Vol 488 / Issue No 7409

THIS WEEK

EDITORIALS

5 SCIENCE & SOCIETY
Unfortunate oversight
Transparency is key when scientists have links with industry

5 RESEARCH FUNDING
Marching orders
To change policy, scientists must be outspoken and proactive

WORLD VIEW

7 **The time is right to confront misconduct**
Colin Macilwain
The problem of scientific fraud is finally being acknowledged

RESEARCH HIGHLIGHTS

8 SELECTIONS FROM THE SCIENTIFIC LITERATURE
'Photoswitch' restores sight / Immunity boost / The 'hunter gatherer' genome / A better battery / Aerosols control monsoon / The high cost of squid sex

SEVEN DAYS

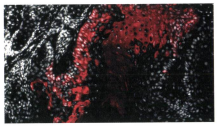
10 THE NEWS IN BRIEF
University of California regents avoid criminal charges / Ebola outbreak in Uganda / Greenland ice sheet suffers massive surface melt / India curbs tourist access to tiger reserves

CAREERS

121 COLUMN
The roots of research misconduct
William Neaves
Mentors have a duty to root out misconduct among their trainees

122 CAREER BRIEFS

NEWS IN FOCUS



13 ONCOLOGY
Cell trackers catch cancer stem cells at work

14 THERAPEUTICS
FDA wins right to regulate stem-cell treatments

15 INFECTIOUS DISEASE
Officials working on strategies to prevent reintroduction of rinderpest

16 PLANETARY SCIENCE
The countdown to Curiosity rover's descent to Mars

18 NUTRITIONAL SUPPLEMENTS
Lawsuit challenges link between telomere length and anti-ageing drug

19 PUBLISHING
Biologists adopt preprint servers

FEATURES

24 SOCIAL SCIENCE
History as science
Historians wary of bid to make their discipline more 'scientific'

FUNDING

Duel to the death

UK research council cuts have physical scientists up in arms PAGE 20

COMMENT

27 SPACE RESEARCH
Let academia lead space science
Daniel N Baker
NASA would save money by investing in the knowledge of university scientists

BOOKS & ARTS

NEUROSCIENCE

Soul purpose

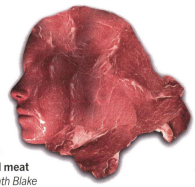
Christof Koch ponders consciousness and our place in the Universe PAGE 29

30 ASTRONOMY
Under African skies
Ivan Semeniuk

31 BOOKS IN BRIEF

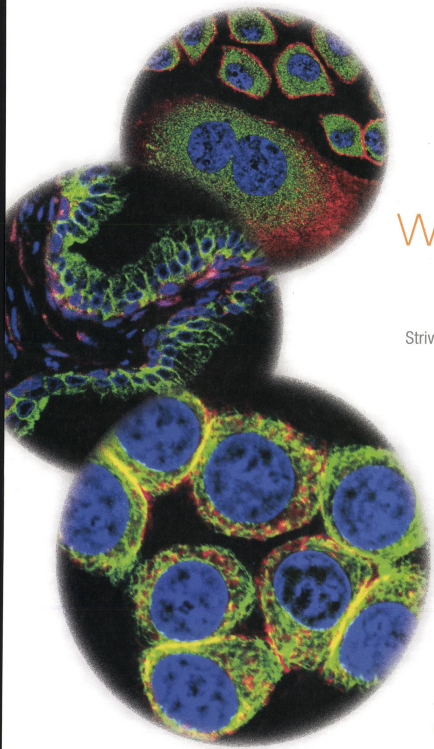
CORRESPONDENCE

32 Antiscience lobby undermining US research funding / Sanitation a priority for China / Family-planning benefits / Bats versus wind turbines



FUTURES

124 **Dead meat**
Polenth Blake



Proteome-wide antibody coverage.

Striving towards an antibody for each human protein.

Unsurpassed development capabilities built over the last ten years allow Abgent to produce more than 5,000 new antibodies per year. Protein A and peptide affinity purification, along with rigorous antibody validation, ensures antibody quality and assay flexibility. The end result is a broad offering of in-house developed, high quality antibodies tailored to your research and drug discovery needs.

Applications: Western Blot, Immunohistochemistry, Immunofluorescence, Flow Cytometry, ELISA, and more.

20µg Evaluation Size Vials
available for most new products.

Visit our website at **www.abgent.com** for more information on Abgent's product and service offerings.

© 2012 ABGENT. All rights reserved. It's all about New is a trademark of ABGENT. For Research Use Only. Not For Use In Diagnostic Procedures. ABG-00104-0612

It's all about New.
Unsurpassed Development Capabilities

ABGENT ANTIBODIES ARE ESSENTIAL TOOLS FOR THE STUDY OF CANCER, NEURODEGENERATIVE DISEASES, METABOLIC SYNDROME, IMMUNOLOGY, APOPTOSIS, AUTOPHAGY, DEVELOPMENT AND MORE.

Visit our website for a complete product listing.

www.abgent.com

药明康德 WuXi AppTec | ABGENT

CONTENTS

2 August 2012 / Vol 488 / Issue No 7409

RESEARCH

NEW ONLINE

33 Papers published this week at nature.com

NEWS & VIEWS

34 PALAEONTOLOGY
An insect to fill the gap
A complete Devonian insect fossil narrows the Hexapoda gap
William A Shear SEE LETTER P.82

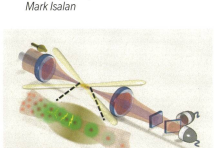
35 EARTH SCIENCE
The balance of the carbon budget
Total global carbon sinks have increased during the past 50 years
Ingeborg Levin SEE LETTER P.70

36 DRUG DISCOVERY
Kill the messenger where it lives
Antisense molecules target a toxic mRNA for destruction in the nucleus
Peter K Todd & Henry L Paulson
SEE LETTER P.111

38 ALZHEIMER'S DISEASE
A protective mutation
Gene variant reduces accumulation of amyloid-β peptides in the brain
Bart De Strooper & Thierry Voet
SEE LETTER P.96

39 QUANTUM OPTICS
Strongly interacting photons
A medium that is transparent to single photons
Thad G Walker SEE LETTER P.57

40 SYSTEMS BIOLOGY
A cell in a computer
Modelling every cellular process in dividing *Mycoplasma* bacteria
Mark Isalan



PHYSICS

Light work

An step towards all-optical quantum information processing. PAGE 57

ON THE COVER

Ground work

Two groups have characterized the root-dwelling bacteria of the classic model plant, *Arabidopsis*. They find that bacterial communities in each root compartment are strongly influenced by soil type, and to a lesser degree by host genotype. PAGES 86 & 91

TIM ROBERTS/GETTY IMAGES

ARTICLES

43 CANCER **Novel mutations target distinct subgroups of medulloblastoma**
G Robinson et al.

49 CANCER **Subgroup-specific structural variation across 1,000 medulloblastoma genomes**
P A Northcott et al.

LETTERS

57 PHYSICS **Quantum nonlinear optics with single photons enabled by strongly interacting atoms**
T Peyronel et al. SEE N&V P.39

61 MATERIALS **Quantum phase transition in a resonant level coupled to interacting leads**
H T Mebrahtu et al.

65 PHYSICS **A Newtonian approach to extraordinarily strong negative refraction**
H Yoon, K Y M Yeung, V Umansky & D Ham

70 ENVIRONMENT **Increase in observed net carbon dioxide uptake by land and oceans during the past 50 years**
A P Ballantyne, C B Alden, J B Miller, P P Tans & J W C White SEE N&V P.35

73 EARTH SCIENCE **Persistent near-tropical warmth on the Antarctic continent during the early Eocene epoch**
J Pross et al.

78 ECOLOGY **Universal species–area and endemics–area relationships at continental scales**
D Storch, P Keil & W Jetz

82 EVOLUTION **A complete insect from the Late Devonian period**
R Garrouste et al. SEE N&V P.34

86 MICROBIOLOGY **Defining the core *Arabidopsis thaliana* root microbiome**
D S Lundberg et al.

91 MICROBIOLOGY **Revealing structure and assembly cues for *Arabidopsis* root-inhabiting bacterial microbiota**
D Bulgarelli et al.

96 CELL BIOLOGY **A mutation in *APP* protects against Alzheimer's disease and age-related cognitive decline**
T Jonsson et al. SEE N&V P.38

100 CANCER **Dissecting the genomic complexity underlying medulloblastoma**
D T W Jones et al.

106 CANCER **Medulloblastoma exome sequencing uncovers subtype-specific somatic mutations**
T J Pugh et al.

111 THERAPEUTICS **Targeting nuclear RNA for *in vivo* correction of myotonic dystrophy**
T M Wheeler et al. SEE N&V P.36

116 GENOMICS **A map of the *cis*-regulatory sequences in the mouse genome**
Y Shen et al.

EVOLUTION

Leg work

A complete six-legged insect bridges a gap in the fossil record. PAGE 82

nature

2 August 2012 / Vol 488 / Issue No. 7409

EDITORIAL

LONDON, MUNICH, WASHINGTON DC, NEW YORK, BOSTON, PARIS, SAN FRANCISCO, TOKYO nature@nature.com

EDITOR-IN-CHIEF: Philip Campbell
EXECUTIVE EDITOR: Nick Campbell
PUBLISHING EXECUTIVE EDITOR: Maxine Clarke **EXECUTIVE EDITOR:** Véronique Kiermer
CHIEF MAGAZINE EDITOR: Tim Appenzeller
THIS WEEK/NEWS/FEATURES: Alison Abbott, David Adam, Ananyo Bhattacharya, Geoff Brumfiel, Declan Butler, Ewen Callaway, Erika Check Hayden, Daniel Cressey, David Cyranoski, Natasha Gilbert, Eric Hand, Heidi Ledford, Corie Lok, Brendan Maher, Emma Marris, Richard Monastersky, Brian Owens, Helen Pearson, Mark Peplow, Eugenie Samuel Reich, Ivan Semeniuk, Quirin Schiermeier, Jeff Tollefson, Richard Van Noorden, Meredith Wadman, M Mitchell Waldrop **COMMENT:** Sara Abdulla, Joanne Baker, Rosalind Cotter, Barbara Kiser, Alison McCook, Lucy Odling-Smee, Sarah Tomlin **NATURE PODCAST/VIDEO:** Adam Rutherford, Thea Cunningham, Geoff Marsh, Kerri Smith, Charlotte Stoddart **NEWS AND VIEWS:** Sadaf Shadan, Ana Lopes, Andrew Mitchinson, Cesar Sanchez, Marian Turner **RESEARCH:** Ritu Dhand, Kari Ziemelis, Francesca Cesari, I-han Chou, Tanguy Chouard, Rosamund Daw, Alex Eccleston, Angela Eggleston, Joshua Finkelstein, Henry Gee, Patrick Goymer, Noah Gray, Marie-Thérèse Heemels, Magdalena Helmer, Karen Howell, Claudia Lupp, Barbara Marte, Andrew Mitchinson, Juliane Mössinger, Deepa Nath, Wayne Peng, Leslie Sage, Magdalena Skipper, Clare Thomas, Maria Trajkovska, John VanDecar, Liesbeth Venema, Ursula Weiss, Michael White **INSIGHTS/REVIEWS/PERSPECTIVES:** Ursula Weiss **CAREERS:** Gene Russo, Karen Kaplan **TECHNOLOGY FEATURES:** Monya Baker **FUTURES:** Henry Gee **SUPPLEMENTS:** Herb Brody, Michelle Gyssyn, Tony Scully **SUBEDITORS:** Colin Sullivan, Sarah Archibald, Nicola Bailey, Anne Blewett, Nicolas Fanget, Isobel Flanagan, Paul Fletcher, Sasha Lewis, Dinah Loon, Louisa Lyons, Anna Novitzky, Yukie Ozawa, Katherine Perry, David Price, Jenny Rooke, David Winters, Anna York **PRODUCTION:** Jenny Henderson, Leonora Dawson-Bowling, Dan Foley, Stewart Fraser, Susan Gray, Simon Gribbin, Hazel Mayhew, James McQuat, Emilia Orviss, Yvonne Strong, Robert Sullivan, Beverta Tarran, Charles Wenz **ART AND DESIGN:** Kelly Buckheit Krause, Wesley Fernandes, Madeline Hutchinson, Barbara Izdebska, Paul Jackman, Jasiek Krzyszto ak, Annthea Lewis, Chris Ryan, Nik Spencer, Claire Welsh, Mark Wood **JAPAN EDITION:** Eiji Matsuda, Mika Ishida, Mitsuyo Utsugi, Sayaka Ayame **PRESS OFFICE:** Ruth Francis, Neda Afsarmanesh, Lisa Boucher, Rachel Twinn, Rebecca Walton **ADMINISTRATION:** Paul Sng, Liz Bankes, Roseann Campbell, Peter Cary, Pfion Cleverley, Dana Coles, Faye Fornaser, Ashlee Gagui, Amy Invernizzi, Caroline McLean, Rina Nozawa, Anastasia Panoutsou, Kenneth Simpson, Diane Yorke

PUBLISHING

LONDON feedback@nature.com

MANAGING DIRECTOR & PUBLISHER: Steven Inchcoombe
ASSOCIATE PUBLISHER: Arash Hejazi
PUBLISHING PROJECT MANAGERS: Claudia Deasy, Christian Manco

TOKYO feedback@natureasia.com

REGIONAL MANAGING DIRECTOR — ASIA-PACIFIC: David Swinbanks
DIRECTOR — ASIA-PACIFIC: Antoine E Bocquet
OPERATIONS DIRECTOR — ASIA-PACIFIC: Hiroshi Minemura

DISPLAY ADVERTISING, SPONSORSHIP & NATUREJOBS

GLOBAL HEAD: Andrew Douglas a.douglas@nature.com

DISPLAY ADVERTISING AND SPONSORSHIP

GLOBAL HEAD: Gerard Preston g.preston@nature.com
ASSISTANT SPONSORSHIP MANAGER: Reya Silao Tel: +44 (0)20 7843 4977 (rsilao@nature.com)

EUROPE/REST OF WORLD display@nature.com Tel: +44 (0)20 7843 4960
GERMANY/SWITZERLAND/AUSTRIA: Sabine Hugi-Fürst Tel: +41 52761 3386
UK/IRELAND/SCANDINAVIA/SPAIN/PORTUGAL: Evelina Rubio Hakansson Tel: +44 (0)20 7843 4079
FRANCE/BELGIUM/THE NETHERLANDS/LUXEMBOURG/ITALY/OTHER EUROPE: Nicola Wright Tel: +44 (0)20 7843 4959

NORTH AMERICA display@natureny.com
NEW ENGLAND/ EASTERN CANADA: Sheila Reardon Tel: +1 617 494 4900
NEW YORK/MID-ATLANTIC/SOUTHEAST: Jim Breault Tel: +1 212 726 9334
MIDWEST: Mike Rossi Tel: +1 212 726 9255
WEST COAST/ WESTERN CANADA: George Lui Tel: +1 415 781 3804
WEST COAST NORTH: Kayla McCutchan Tel: +1 415 403 9038

ASIA-PACIFIC display@natureasia.com
ASIA-PACIFIC HEAD: Kate Yoneyama
JAPAN: Ken Mikami Tel: +81 3 3267 8765
GREATER CHINA/SINGAPORE: Gloria To Tel: +852 2811 7191
INDIA: Vikas Chawla Tel: +91 124 288 1057

SPONSORSHIP

BUSINESS DEVELOPMENT MANAGER: David Bagshaw Tel: +1 212 726 9215
BUSINESS DEVELOPMENT MANAGER: Patrick Murphy Tel: +1 617 475 9216

NATUREJOBS naturejobs@nature.com
EUROPEAN HEAD OFFICE, LONDON: Tel: +44 (0)20 7843 4961
EUROPEAN SALES MANAGER: Nils Moeller (n.moeller@nature.com) **BUSINESS DEVELOPMENT MANAGER, MIDDLE EAST:** Jon Giuliani (j.giuliani@nature.com) **ADVERTISING PRODUCTION MANAGER:** Stephen Russell **NATUREJOBS ONLINE PRODUCTION:** Dennis Chu
US HEAD OFFICE, NEW YORK: naturejobs@natureny.com **US SALES MANAGER:** Ken Finnegan
JAPAN HEAD OFFICE, TOKYO: Tel: +81 3 3267 8765 **ASIA-PACIFIC HEAD:** Kate Yoneyama
ASIA-PACIFIC SALES & BUSINESS DEVELOPMENT MANAGER: Yuki Fujiwara **BUSINESS DEVELOPMENT MANAGER, GREATER CHINA, SINGAPORE:** Gloria To Tel: +852 2811 7191
INDIA: Vikas Chawla Tel: +91 1242881057 (v.chawla@nature.com)

MARKETING & SUBSCRIPTIONS

USA/CANADA/LATIN AMERICA subscriptions@us.nature.com
Tel: (USA/Canada)+1 866 363 7860; (outside USA/Canada) +1 212 726 9223
MARKETING: Sara Girard **FULFILMENT:** Karen Dawson

JAPAN/CHINA/KOREA subscriptions@natureasia.com
Tel: +81 3 3267 8751 **MARKETING:** Sachiyo Ikeda **FULFILMENT:** Masayo Tanaka

EUROPE/REST OF WORLD subscriptions@nature.com
Tel: +44 (0)1256 329242
MARKETING: Elena Woodstock, Marcus Parker, Hannah Phipps **FULFILMENT:** Jean Burrows

INDIA npgindia@nature.com
Tel: +91 124 2881053/54
HEAD OF BUSINESS DEVELOPMENT, INDIA: Debashish Brahmachari **MARKETING:** Harpal Singh Gill

ANNUAL SUBSCRIPTIONS (UK/Europe, including post and packing)
FULL RATE: £1,400
PERSONAL RATES: UK £135; Europe €209; Rest of World £210
STUDENT RATES: UK £79; Europe €122; Rest of World £125
Individual rates apply to payments by personal cheque or credit card. Back issues: £20.00.

ANNUAL SUBSCRIPTIONS (USA, including post and packing)
INSTITUTIONAL/CORPORATE RATE: $2,730 ; **PERSONAL RATE:** $199; **STUDENT RATE:** $99;
POSTDOC RATE: $119
Printed in USA. Individual rates available only to subscribers paying by personal check or credit card. Orders for student/postdoc subscriptions must be accompanied by a copy of student ID. Rates apply to USA, Canada, Mexico/Central & South America. Add 7% GST tax in Canada (Canadian GST number 140911595). **BACK ISSUES:** US$20.00.

ANNUAL SUBSCRIPTIONS (Asia, including post and packing)
JAPAN: 年間購読料金：*Nature* のみ 437,800 円 +21,890 円 (消費税) =459,690 円 (合計)
Nature + *Nature Digest* 466,550 円 +23,328 円 (消費税) =489,878 (合計)
お問い合わせ 日本出版貿易株式会社 国内営業 1 部
〒101-0064 東京都千代田区猿楽町1-2-1 Tel: 03-3292-3767 Fax: 03-3292-8766
個人/学生購読について
NPG ネイチャー アジア パシフィック フルフィルメント & カスタマーサービス
〒162-0843 東京都新宿区市谷田町2-37 ネイチャービル Tel: 03-3267-8759 Fax: 03-3267-8752
CHINA: www.natureasia.com/cn/subscribe; **KOREA:** www.natureasia.com/korea/subscribe

NATURE ONLINE WWW.NATURE.COM/NATURE

CHIEF TECHNOLOGY OFFICER: Howard Ratner **ASSOCIATE DIRECTOR, NATURE.COM:** Daniel Pollock
PLATFORM DEVELOPMENT: Jeremy Macdonald **WEB PRODUCTION:** Alexander Thurrell, Deborah Anthony **WEB DESIGN:** Juan-Carlos Sobrino, Luke Stavenhagen **PLATFORM INTERFACES:** Amanda Ward

SITE LICENCES, FULFILMENT & CUSTOMER SERVICES

SITE LICENCES: npg.nature.com/libraries feedback@nature.com
CHIEF OPERATIONS OFFICER: John Carroll **ASSOCIATE DIRECTOR OF OPERATIONS:** Dominic Pettit
CUSTOMER SERVICE: Liz Davidson **ORDER PROCESSING OPERATIONS:** Steve Poulter

NATURE PUBLISHING GROUP OFFICES

LONDON: The Macmillan Building, 4 Crinan Street, London N1 9XW Tel: +44 (0)20 7833 4000 Fax: +44 (0)20 7843 4596/7

BASINGSTOKE: Nature Publishing Group, Subscriptions, Brunel Road, Basingstoke, Hants RG21 6XS, UK Tel: +44 (0)1256 329242

WASHINGTON DC: 968 National Press Building, 529 14th St NW, Washington DC 20045-1938 Tel: +1 202 737 2355

NEW YORK: 75 Varick St, 9th Floor, New York, NY 10013-1917 Tel: +1 212 726 9200

BOSTON: 25 First Street, Suite 104, Cambridge, MA 02141 Tel: +1 617 475 9275

SAN FRANCISCO: 225 Bush Street, Suite 1453, San Francisco, CA 94104 Tel: +1 415 403 9027

MUNICH: Nymphenburger Strasse 14, D-80335 München Tel: +49 89 549057-13

PARIS: Moulin de Malicorne, Sainte Maure de Touraine, Paris 37800 Tel: +33 2 47 35 72 15

TOKYO: Chiyoda Building, 2-37 Ichigayatamachi, Shinjuku-ku, Tokyo 162-0843 Tel: +81 3 3267 8751

INDIA: Nature Publishing Group, 3A, 4th Floor, DLF Corporate Park, Gurgaon 122002 Tel: +91 124 2881053/54

Nature® (ISSN 0028-0836) is published by Nature Publishing Group, a division of Macmillan Publishers Ltd (The Macmillan Building, 4 Crinan Street, London N19XW). Registered as a newspaper at the British Post Office. North and South American orders to subscriptions@nature.com (Tel: +1 866 363 7860). Other orders to subscriptions@nature.com (Tel. +44 (0)1256 329242; Fax +44 (0)1256 812358). Authorization to photocopy material for internal or personal use, or internal or personal use of specific clients, is granted by *Nature* to libraries and others registered with the Copyright Clearance Center (CCC) Transactional Reporting Service, provided the relevant copyright fee is paid direct to CCC, 222 Rosewood Drive, Danvers MA 01923, USA. Identification code for *Nature*: 0028-0836/03. CPC PUB AGREEMENT #40032744. In the US, *Nature* (ISSN 0028-0836) is published weekly on Thursday, except the last week in December by Nature Publishing Group, 75 Varick St, 9th Fl, New York NY 10013-1917, USA. US Periodicals postage paid at New York NY, and additional mailing post offices. US POSTMASTER: send address changes to *Nature*, Subscription Dept, 75 Varick St, 9th Fl, New York NY 10013-1917, USA. Published in Japan by NPG Nature Asia-Pacific, Chiyoda Building, 2-37 Ichigayatamachi, Shinjuku-ku, Tokyo 162-0843, Japan. ©

2012 Macmillan Publishers Limited. All rights reserved.

THIS WEEK

EDITORIALS

WORLD VIEW The time is right to crack down on research misconduct **p.7**

RECHARGE New lithium–air batteries come with extended life **p.8**

ADAPTATION Early peoples bounced back from climate change **p.9**

Unfortunate oversight

Scientists must remember that however irrelevant their involvement in industry might seem to them, others will see it differently — only full disclosure will avert the taint of scandal.

Hydraulic fracturing, or 'fracking', a technology that revolutionized the natural-gas industry, has been surrounded by controversy in recent years. So, when environmental experts at the University of Texas at Austin produced a report in February that gave the technique a fairly clean bill of health, they received widespread news coverage, including in the pages of *Nature* (see *Nature* **482,** 445; 2012). The study was billed as an independent analysis. Yet last week it emerged that its lead author is a well-paid board member of an energy company that is actively involved in fracking.

The failure to declare this involvement was an unfortunate mistake to make, not least because the man who made it is a respected senior scientist who headed the US Geological Survey under US presidents Bill Clinton and George W. Bush — and is therefore experienced enough to understand the role that politics and perception have in sensitive issues such as energy development. Yet Charles 'Chip' Groat, associate director of the University of Texas at Austin Energy Institute, failed to disclose that he holds a significant number of shares in the Houston-based Plains Exploration & Production Company, and that he earned more than US$400,000 from the company last year. In a 23 July statement to Bloomberg news, he said that disclosing his position on the board "would not have served any meaningful purpose relevant to this study".

Groat says that his position on the board did not affect the outcome of the study and that he did not interfere with the findings of his colleagues. The study found no evidence of groundwater contamination from fracking, which pumps fluid into the ground at high pressure to fracture geological formations and release natural gas or oil. The technology has been in use for decades, and practised properly, the report suggested, it is safe and poses little risk to the environment.

This over-arching conclusion seems reasonable in view of what we know today, although scientists continue to sift through contradictory evidence. And Groat's explanation of his role also sounds plausible — but that is all the more reason for him to have openly disclosed his ties to the industry.

After the link was revealed by the Public Accountability Initiative, a non-profit watchdog in Buffalo, New York, university officials announced plans to review the study. But even if the review exonerates the panel and endorses its findings, it is unlikely to remove the taint of scandal. Rather than cutting through the confusion on fracking, the report is likely to contribute to it.

Experts in many fields bounce between academia, government and industry during their careers. Universities could not exclude people who have industry connections from their ranks, nor would they want to. The same goes for government. There is also nothing inherently wrong with universities accepting donations from industry to conduct studies, as long as the proper protections are put in place. The key is transparency, because that is the basis for trust between institutions and the wider public, which is especially important when people are buffeted by confusing, contradictory and inflammatory information. What the public needs, and what scientists must deliver, is reliable information that is honest about both its methods and its inevitable biases. What it needs is full disclosure. ■

Marching orders

Scientists unhappy with policy are right to take to the streets.

The mock funeral — an idea so good that scientists had it twice. Last month, about 2,000 researchers marched on Parliament Hill in Ottawa, carrying a coffin that signified, they said, the "death of evidence". The scientists were protesting against a series of cuts by Canadian Prime Minister Stephen Harper's conservative government that they believed threatened basic research and undermined expert advice in areas such as environmental policy. And in May, physical scientists drove a horse-drawn Victorian hearse to the British Prime Minister's residence in Downing Street, London, this time to mark the demise of UK science.

The Downing Street stunt was to protest against moves made by the main public funder of UK physical-sciences research, the Engineering and Physical Sciences Research Council (EPSRC), to cut the number of proposals it receives and to prioritize research that addresses national priorities or comes with economic spin-offs (see page 20).

Echoing their Canadian counterparts, the scientists argued that the changes would endanger blue-skies research in chemistry, physics and mathematics. But unlike Canada's protests, the UK campaign has yet to win support from the wider scientific community.

In part, that is because the campaign targets a single, specific funder and so is not seen as relevant to UK science as a whole. Some researchers have dismissed the coffin parade as an overreaction to a spat between a few disenfranchised scientists and the EPSRC. Others worry that a public protest that exposes disunity in the ranks of science at a time of economic chaos could result in cuts to the science budget.

Perhaps, but if it is an isolated spat, then why did people with little personal stake in the EPSRC's policies join in the protests? And the calls by dissenters to close ranks — to keep calm and to carry on — ignore the fact that science funding is a political question. To make a point in a political arena, scientists must stand up and be counted. ■

CUSTOM MONOCLONAL ANTIBODIES
Rat mAb ◦ Mouse mAb ◦ Ascites

CUSTOM PROTEIN EXPRESSION
HEK293 ◦ Baculovirus ◦ *E. coli*

Superior Quality • Unriveled Expertise
www.promab.com
1.866.339.0871 | info@promab.com

ProMab Biotechnologies, Inc.

nature**jobs**

Read our Postdoc articles and news in the **News and Comment** section of **naturejobs.com**

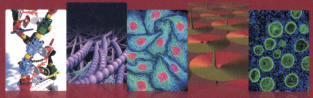
- Helpful advice
- Real life experiences
- Follow the highs and lows of postdoc life

nature**conferences**

Unique events committed to scientific excellence

Sign up for natureconferences e-alerts at www.nature.com/natureconferences

Receive updates on
- Important deadlines
- Newly added conferences
- Highlights from past conferences

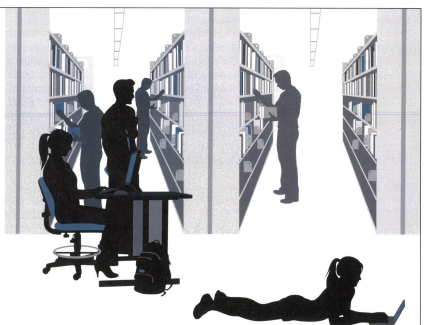


www.nature.com/natureconferences

nature publishing group npg

SCIENTIFIC AMERICAN™

Gain complete online access to the latest in science, technology and medicine **at your institution**



Recommend site license access to your library today!
www.nature.com/scientificamerican

™ Scientific American, Inc.

nature publishing group npg

The time is right to confront misconduct

After a generation of denial, research leaders are finally treating scientific fraud with the seriousness it deserves, says **Colin Macilwain**.

One problem with having worked as a journalist for a long time is that every story comes with a feeling of déjà vu. You keep thinking: I've been here before. So it is refreshing to report one issue where something has actually changed: the vexed and perennial problem of research misconduct, which scientific leaders are finally taking seriously. Talking to several leaders in recent weeks, I have found that their mood has hardened — and not before time.

For too long, scientists' instinctive defensiveness has produced general denial that misconduct constitutes a serious problem.

I arrived in Washington DC to work for *Nature* in 1993, in the aftermath of congressional hearings into allegations of misconduct involving a paper by biologists David Baltimore and Thereza Imanishi-Kari at the Massachusetts Institute of Technology in Cambridge. The researchers were correctly found innocent. But the case led an independent commission chaired by reproductive biologist Kenneth Ryan to call for a much more rigorous approach to the investigation of misconduct.

Ryan was shot down in flames by scientific officials and his recommendations were ignored. They were delivered to the US Department of Health and Human Services, which kicked them upstairs to the White House. The administration of then-president Bill Clinton sat on the findings until 2000, when it issued a bland federal misconduct decree. And that was in the United States — the world's dominant scientific power and the one that had done the most to address misconduct.

Countermeasures elsewhere have been even feebler. In Germany, for example, no university had an integrity officer until 2011, and it is still difficult for institutions there to sanction proven fraudsters. Some judges consider academic freedom of expression to be paramount — and say that it would be violated if a university were to request scientists to retract a paper.

Worldwide, however, research integrity is now very much in the spotlight. Prominent cases in the United Kingdom, South Korea, the Netherlands and Canada in recent years have each had a disturbing and powerful impact in their respective locales.

Considerable hard data have emerged on the scale of misconduct. A metastudy (D. Fanelli *PLoS ONE* **4,** e5738; 2009) and a detailed screening of all images in papers accepted by *The Journal of Cell Biology* (M. Rossner *The Scientist* **20** (3), 24; 2006) each suggest that roughly 1% of published papers are fraudulent. That would be about 20,000 papers worldwide each year.

At the time of the Baltimore case, it was widely argued that research misconduct was insignificantly rare — and irrelevant to the progress of science, which would self-correct. Few senior scientists now believe that. They know that misconduct exists and that, unchecked, it can undermine public regard for science and scientists.

Two major studies to be released in the next year reflect this shift in attitude. Significantly, they have been instigated by leading scientists. One study, by the InterAcademy Council, is looking at international aspects of misconduct. Sharp disparities in investigative procedures — and the lack of any such procedures, or responsible officials, at many institutions outside the United States — are problematic, given that an increasing proportion of research involves collaborators from more than one country.

Robbert Dijkgraaf, co-chairman of the InterAcademy Council, is one of the people leading the study. He hopes that, when its findings are released this year, governments and research agencies around the world will use them as a template to improve training and enforcement of good research conduct.

CURRENT SCIENTIFIC LEADERS HAVE THE OPPORTUNITY TO TAKE THE INITIATIVE AND STAMP DOWN ON FRAUD.

The second study, by the US National Academy of Sciences, will report in 2013. It is likely to call for far-reaching changes in how US agencies define and police misconduct. Since the 2000 decree, agencies have regarded only 'falsification, fabrication and plagiarism' as misconduct: the academy may call for this definition to be widened in line with an emerging global consensus to include most other sorts of unethical behaviour, such as falsely attributed authorship.

Last December, for example, Canada established a Tri-Agency Framework for the Responsible Conduct of Research at its main funding agencies. The framework oversees publicly and privately funded research and has a secretariat to support university misconduct investigations.

Britain is also finally taking some faltering steps to address the issue. In July, universities adopted a voluntary concordat that obliges them to investigate misconduct allegations. Some research leaders want to leave it at that but others, led by Michael Rawlins, chairman of the UK National Institute for Health and Clinical Excellence, demand further action to ensure that cases are properly investigated.

Current scientific leaders have the opportunity to take the initiative and stamp down hard on fraud. Next year's National Academy study won't use language as divisive as Ryan's, but it could usher in a more consistent US system to handle misconduct, which could percolate around the globe. The international report will help governments and agencies to pursue miscreants across borders. Together, the studies represent a historic opportunity to deal with what is, perhaps, the single most potent threat to science's prestige. ■

➲ **NATURE.COM**
Discuss this article online at:
go.nature.com/6knitx

Colin Macilwain *writes about science policy from Edinburgh, UK. e-mail: cfmworldview@gmail.com*

RESEARCH HIGHLIGHTS

Selections from the scientific literature

NEUROSCIENCE

Blind mice can sense light

A small molecule restores light sensitivity to blind mice when it is injected into their eyes.

Richard Kramer at the University of California, Berkeley, and his colleagues studied a mouse model of retinitis pigmentosa — a form of blindness in which light-sensing rod and cone cells in the retina degenerate. The researchers applied the molecule, AAQ, to retinas isolated from the mice and found that it triggered retinal ganglion cells — most of which are normally light-insensitive — to increase their firing rate in response to light. Other work has suggested that the molecule functions by blocking potassium ion channels in the membranes of neurons, boosting their excitability. In behavioural tests, blind mice treated with AAQ showed signs of light sensitivity.

The use of this and related molecules could restore vision less invasively than other proposed methods, the researchers say.

Neuron 75, **271–282 (2012)**

MICROBIOLOGY

Skin bacteria boost immunity

Microbes living in mammalian guts have an important role in intestinal immunity and it seems that those living on the skin are similarly crucial for tuning immune responses to skin pathogens.

Yasmine Belkaid at the US National Institute of Allergy and Infectious Diseases in Bethesda, Maryland, and her colleagues compared mice with microbes on their skin with germ-free mice raised in aseptic conditions. T cells, a subset of immune cells, produced fewer immune-stimulating molecules in germ-free animals than in control mice. When infected with a skin parasite, the germ-free mice developed a greater number of parasites per skin lesion than the controls, and also showed impaired T-cell responses. Populating the skin with a skin bacterium restored immunity to the germ-free animals.

Science **http://dx.doi.org/10.1126/science.1225152 (2012)**

For a longer story on this research, see go.nature.com/8ahyc3

N. PAVITT/JAI/CORBIS

GENOMICS

Hunter-gatherer genes

Three African populations that rely mainly on hunting and gathering possess a trove of previously unrecorded genetic diversity.

Sarah Tishkoff at the University of Pennsylvania in Philadelphia and her team sequenced the full genomes of five individuals from each of three populations: Cameroonian Pygmies, and the Hadza (pictured) and Sandawe people from Tanzania. The researchers' trawl uncovered 13.4 million variants — more than 3 million of which have never been seen before.

Genes involved in immunity, metabolism, taste, smell and reproduction seem to have evolved since the different populations split — a sign of adaptation to local environments. In the Pygmies, recent changes in genes involved in the function of the pituitary gland, which secretes growth and other hormones, could explain their short stature. All the hunter-gatherers sampled showed signatures of gene flow from now-extinct human species. This has been seen before mainly in non-African populations, supporting the idea that breeding between various human species occurred regularly.

Cell **http://dx.doi.org/10.1016/j.cell.2012.07.009 (2012)**

For a longer story on this research, see go.nature.com/ss7rzr

ELECTROCHEMISTRY

Rechargeable Li–air battery

Lithium–air batteries promise to greatly exceed the energy-storage capacity of conventional lithium-ion batteries and a study shows that they can retain 95% of their capacity even after 100 recharges.

Current is generated in lithium–air batteries when lithium ions from the anode react with oxygen from the air — rather than with a limited volume of oxidizing agent, as in conventional batteries. Peter Bruce and his colleagues at the University of St Andrews, UK, created a lithium–air battery using an electrolyte of dimethylsulphoxide, through which the lithium ions flow, and a porous gold cathode where oxygen is reduced before it reacts with the lithium ions. These materials seem to prevent the side-reactions that have quickly degraded the performance of previous lithium–air batteries.

Science **http://dx.doi.org/10.1126/science.1223985 (2012)**

ANIMAL BEHAVIOUR

Sex is costly for squid

For a squid, mating can take up to three hours, and the resulting energy losses could put the animal at a disadvantage around predators and reduce

M. NORMAN/MUSEUM VICTORIA

foraging opportunities.

Amanda Franklin and her colleagues at the University of Melbourne in Victoria, Australia, collected wild dumpling squid (*Euprymna tasmanica*; **mating male and female pictured**). They tested the creatures' swimming endurance in a tank with a constant current, before and after the squid mated. Mating halved the time taken for males and females to become exhausted, but both regained their energy within 30 minutes of copulation.

Knowing this cost could contribute to a better understanding of the evolution of reproductive behaviours, such as promiscuity, in squid, the researchers say.

Biol. Lett. **http://dx.doi.org/10.1098/rsbl.2012.0556 (2012)**

NEUROSCIENCE

Light control in monkey brains

Using a technique that makes it possible to control the activity of specific engineered neurons with light, neuroscientists have modified behaviour in primates.

Optogenetics can help researchers to figure out the role of individual neurons, but it has previously been used to control behaviour only in rodents and invertebrates. Wim Vanduffel at Harvard Medical School in Boston, Massachusetts, and his colleagues genetically modified neurons in the premotor and prefrontal cortex brain regions of two rhesus monkeys, so that the neurons fired when blue light was delivered into the brain by an optical cable. Stimulating these neurons as the monkeys performed an eye-movement task changed the latencies of eye movements in both animals. Functional magnetic resonance imaging revealed that the stimulation induced distinct patterns of brain activity during the task.

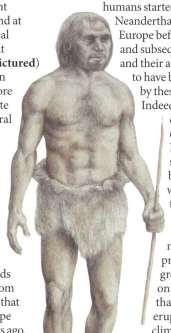
Curr. Biol. **http://dx.doi.org/10.1016/j.cub.2012.07.023 (2012)**

PALAEOANTHROPOLOGY

Resilient to natural disasters

A study of ancient volcanic ash found at key archaeological sites suggests that Neanderthals (**pictured**) and early modern humans were more resilient to climate change and natural disasters than is often assumed.

John Lowe at the Royal Holloway University of London in Egham, UK, and his colleagues analysed microscopic shards of volcanic ash from a major eruption that occurred in Europe some 40,000 years ago. The volcano spewed so much climate-cooling ash that the event probably created winter-like conditions. Because the researchers found the ash at several archaeological sites in Europe and North Africa, they were able to link events in Neanderthal and human evolution with the timing of climatic changes. Early modern humans started to displace Neanderthals from parts of Europe before the eruption and subsequent cooling, and their activities appear to have been unaffected by these events. Indeed, in parts of central and eastern Europe, Neanderthals seem to have become extinct well before the eruption occurred. Early modern humans probably placed greater pressure on Neanderthals than did volcanic eruptions or climate change, the researchers suggest.



Proc. Natl Acad. Sci. USA **http://dx.doi.org/10.1073/pnas.1204579109 (2012)**

COMMUNITY CHOICE

The most viewed papers in science

CLIMATE MODELLING

Aerosols keep down monsoon rain

HIGHLY READ on www.agu.org the week of 23 July

Tiny airborne particles called atmospheric aerosols tend to reduce summer monsoon rainfall over most of South Asia.

Dilip Ganguly and his colleagues at the Pacific Northwest National Laboratory in Richland, Washington, used a simplified atmosphere–ocean model to simulate the effects of changes in the levels and composition of atmospheric aerosols — from local and distant sources — on South Asia's mean monsoon rainfall. Increased local emissions of aerosols such as black carbon — which absorbs sunlight and produces a warming effect that tends to reduce cloud cover — weakened the monsoon rains in most of South Asia. Aerosols from outside Asia also contributed to the overall reduction in rainfall.

Only over northwest India, where aerosol emissions from local forest and grass fires are thought to be decreasing, did the mean summer monsoon rainfall increase.

J. Geophys. Res. **http://dx.doi.org/10.1029/2012JD017508 (2012)**

GENE THERAPY

Gene fix repairs hearing

Gene therapy has restored hearing for up to 18 months in mice that were born deaf.

The animals are missing the gene that encodes the protein VGLUT3. Lack of VGLUT3 renders inner hair cells of the ear's cochlea incapable of sending electrical signals to the brain. Lawrence Lustig at the University of California, San Francisco, and his team used a virus to deliver the *Vglut3* gene into the cochleas of these mice. After one week, the researchers detected auditory responses in the creatures' brains, and within two weeks, the animals showed an increased startle response to sound.

The results could bode well for humans, the researchers suggest, because VGLUT3 is also associated with a rare form of human deafness.

Neuron **75, 283–293 (2012)**

BRIDGEMAN ART LIBRARY/GETTY

NATURE.COM

For the latest research published by *Nature* visit:

www.nature.com/latestresearch

SEVEN DAYS *The news in brief*

POLICY

Stem-cell ruling

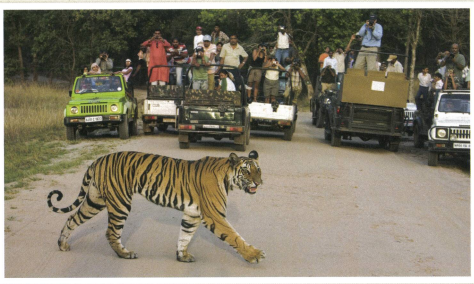
The US Food and Drug Administration (FDA) has now been given legal backing for its attempts to regulate a US clinic that offers therapies involving a patient's processed stem cells. Such treatments can now be classified as drugs, a US District Court in Washington DC ruled on 23 July, in relation to the FDA's injunction against Regenerative Sciences, a stem-cell clinic in Broomfield, Colorado. The ruling could pave the way for the agency to regulate other stem-cell clinics. See page 14 for more.

Data exemption

A cross-party group of politicians has recommended that England's laws on freedom of information be modified to protect universities from having to release data prematurely. The nation's universities have complained that the Freedom of Information Act might be used to force the release of research findings and data before they are ready for publication. On 26 July, Parliament's Justice Select Committee agreed, saying that the existing 'pre-publication exemption' section of the act should be amended. See go.nature.com/cf1fia for more.

Lab-death charges

A landmark criminal prosecution over an accident in a US academic laboratory reached a partial conclusion on 27 July. In a deal that saw criminal charges dropped, the regents of the University of California accepted responsibility for laboratory conditions three-and-a-half years ago, when 23-year-old Sheharbano Sangji died in a lab fire at the University of California, Los Angeles. The regents also agreed to put in place stringent safety measures and to set up a US$500,000 environmental-law scholarship in Sangji's name. But charges remain against Sangji's supervisor, the organic chemist Patrick Harran; his case has been postponed until 5 September. See go.nature.com/hmoden for more.



T. ALLOFS/MINDEN PICTURES/FLPA

India curbs tiger tourism

India's Supreme Court has placed an interim ban on tourists visiting central parts of the country's 40 or so tiger reserves, to protect the dwindling population of the endangered big cats. The 24 July ruling — which the court will re-examine on 22 August — still allows tourists into fringe areas of reserves ('buffer zones'). Park managers said that the ruling would devastate tourism in some national reserves, but that other parks would hardly be affected, because they already keep core areas off-limits. The order came in response to a petition from conservationist Ajay Dubey at the non-governmental organization Prayatna in Bhopal. According to a 2010 census, India is home to about 1,700 wild tigers — more than half of the world's total.

Industry ties

Doubt has been cast on a supposedly independent study into the risks of fracking (the pumping of high-pressure fluids into shale to force out natural gas) after its lead author confirmed last week that he is on the board of directors for an energy company actively involved in the practice, a position that earned him more than US$400,000 last year. Charles Groat, of the University of Texas at Austin, did not disclose his industry ties when the report (see go.nature.com/sopiwm) was released in February. The university says that it is reviewing the study. See page 5 for more.

EVENTS

Ebola outbreak

The first widespread outbreak of Ebola haemorrhagic fever since 2009 has killed 14 people in the Kibaale district of western Uganda, the World Health Organization said on 29 July. Twenty cases have been reported since the beginning of July, but the presence of ebolavirus was not officially confirmed until last week. After the virus spread to the capital, Kampala, Ugandan President Yoweri Museveni told people to avoid physical contact. According to the Uganda Virus Research Institute in Entebbe, the outbreak involves the Sudan subtype of the virus, which in a 2000–01 Ugandan outbreak killed 224 people — 53% of identified cases.

Greenland melt

Satellite observations revealed massive surface melting across the Greenland ice sheet last month as a dome of unusually hot air settled over the region, NASA scientists announced on 24 July. Between 8 and 12 July, the area subject to melting increased from 40% to 97% of the ice sheet — an extent unprecedented in three decades of space observations. The previous record was 55%. But

POLARIS/EYEVINE

the event falls within the realm of natural variability: ice-core records suggest that extreme melting occurs roughly once every 150 years, with the most recent event in 1889.

RESEARCH

Warming redux

The Berkeley Earth Surface Temperature (BEST) study released the second part of its independent analysis of the global land-surface-temperature record on 29 July. The findings — that the planet has warmed over the past 250 years owing to human influence — are not news to climate scientists. But team leader Richard Muller, a physicist at the University of California, Berkeley, is being criticized by researchers for publicizing the BEST results before they have been peer reviewed. The BEST team has not yet published any of its findings in journals, despite posting its first results online last October. See go.nature.com/euvydr for more.

FUNDING

Antarctica upgrade

US research facilities in Antarctica, such as the *Polar Star* icebreaker (**pictured**), need an overhaul, says a report to the US National Science Foundation, released on 23 July. To pay for the upgrade,

the authors recommend that the roughly US$300-million budget for the US Antarctic Program (USAP) be increased by 6%, and that the programme divert 6% of its planned science spending to infrastructure. The USAP devotes nine times more person days in Antarctica to logistics efforts than it does to actual research, and after the upgrade, the balance should tilt more towards research, the authors add. See go.nature.com/dvb9y9 for more.

PEOPLE

Nuclear safety

Japan's new nuclear regulatory commission will probably be headed by radiation physicist Shunichi Tanaka. On 26 July, a parliamentary committee (covering both lower and upper houses) proposed Tanaka as head of the commission, which will be launched in September and will be affiliated with the environment ministry. But some observers objected to Tanaka, a former deputy chair of the cabinet's Japan Atomic Energy Commission, accusing him of being too close to the nuclear industry and playing down the health risks from last year's disaster at the Fukushima Daiichi nuclear plant.

Museum head

The Smithsonian Institution's National Museum of Natural History in Washington DC has announced its next director: Kirk Johnson, a geologist who specializes in plant fossils from the Cretaceous period. Johnson currently serves as vice-president of research and collections and as chief curator at the Denver Museum of Nature & Science in Colorado. He will take on the Smithsonian museum's US$68-million budget and 126 million artefacts and specimens on 29 October. Johnson replaces Cristián Samper, who stepped down on 23 January to head the Wildlife Conservation Society in New York.

Misconduct charge

A once high-flying Danish neuroscientist, Milena Penkowa, is suspected of "potentially intentional misconduct" involving 15 research papers, according to a leaked report from an international committee investigating her case. The report — published on 25 July by Danish newspaper *BT* but due to be released officially in August — had been requested by the University of Copenhagen, Penkowa's former employer, in February 2011. Penkowa had already resigned and been sentenced for embezzling money from the Danish Society for Neuroscience. Two of her papers have been officially retracted, and a report from the Danish Committee on Scientific Dishonesty is also expected later in summer. See go.nature.com/eakrbd for more.

Physics millions

A lucrative prize for fundamental physics was launched on 31 July, with nine researchers each receiving US$3 million. The prize is sponsored by Yuri Milner, a Russian billionaire who once studied for a physics PhD. Milner chose the first winners: Alan Guth, Andrei Linde, Nima Arkani-Hamed, Juan Maldacena, Nathan Seiberg, Edward Witten, Alexei Kitaev, Maxim Kontsevich and Ashoke Sen, who will, in turn, select future winners. The prize will be announced annually and is accompanied by an ad hoc award for 'exceptional cases' and a $100,000 prize for promising junior researchers. See go.nature.com/mwaays for more.

COMING UP

6 AUGUST
NASA tries to land its Curiosity rover on Mars. See page 16 for more.
mars.jpl.nasa.gov/msl

5–10 AUGUST
The Ecological Society of America meets in Portland, Oregon, to discuss preserving, utilizing and sustaining Earth's ecosystems.
www.esa.org/portland

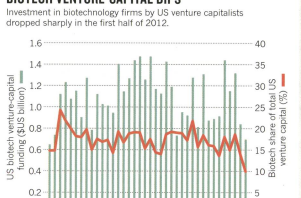
TREND WATCH

US venture capitalists seem to be avoiding the biotechnology sector in favour of information technology, according to numbers released on 20 July in the US National Venture Capital Association's 'MoneyTree' report. Investments in biotech firms dropped to a 9-year low of around US$700 million in the second quarter of 2012. And they accounted for fewer than 10% of all venture-capital deals in that quarter — a much lower proportion than usual (see chart).

SOURCE: THOMSON REUTERS/ WWW.PWCMONEYTREE.COM

BIOTECH VENTURE-CAPITAL DIPS

Investment in biotechnology firms by US venture capitalists dropped sharply in the first half of 2012.



NATURE.COM
For daily news updates see:
www.nature.com/news

nature

Nature Awards for Mentoring in Science – The Nordic Countries

For its annual awards for science mentorship in 2012, *Nature* invites nominations of outstanding scientific mentors in the Nordic countries – Denmark, Finland, Iceland, Norway, Sweden and Greenland. Two prizes of €10,000 (or the equivalent local currency) will be awarded, one for a mid-career mentor and one for life-time achievement in mentoring.

Nominations are now open. The closing date for entries is 13 August 2012.

Nominate your mentor today!

nature.com/nature/mentoringawards/nordiccountrie

nature publishing group npg

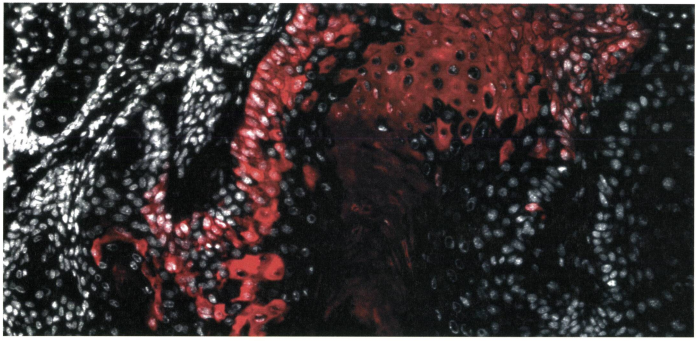
NEWS IN FOCUS

AGRICULTURE Keeping a cattle plague in its grave **p.15**

SPACE The Mars rover's nail-biting descent to the red planet **p.16**

PUBLISHING Biologists flirt with physics-style prepublication **p.19**

SOCIAL SCIENCE Can maths make sense of the cycles of history? **p.24**



G. DRESSEINS

For the first time, researchers can trace cell lineage within a growing tumour. In this skin tumour, the cells labelled red all arose from a single stem cell.

ONCOLOGY

Cancer stem cells tracked

The master builders that underlie tumour growth may inform treatment strategies.

BY MONYA BAKER

Cancer researchers can sequence tumour cells' genomes, scan them for strange gene activity, profile their contents for telltale proteins and study their growth in laboratory dishes. What they have not been able to do is track errant cells doing what is more relevant to patients: forming tumours. Now three groups studying tumours in mice have done exactly that[1–3]. Their results support the ideas that a small subset of cells drives tumour growth and that curing cancer may require those cells to be eliminated.

It is too soon to know whether these results — obtained for tumours of the brain, the gut and the skin — will apply to other cancers, says Luis Parada at the University of Texas Southwestern Medical Center in Dallas, who led the brain study[2]. But if they do, he says, "there is going to be a paradigm shift in the way that chemotherapy efficacy is evaluated and how therapeutics are developed". Instead of testing whether a therapy shrinks a tumour, for instance, researchers would assess whether it kills the right sorts of cell.

Underlying this scenario is the compelling but controversial hypothesis that many tumours are fuelled by 'cancer stem cells' that produce the other types of cancer cell, just as ordinary stem cells produce normal tissues. Previous studies have tested this idea by sorting cells from a cancer biopsy into subsets on the basis of factors such as cell-surface markers, and injecting them into laboratory mice. In principle, those cells that generate new tumours are the cancer stem cells. But sceptics point out that transplantation removes cells from their natural environment and may change their behaviour. "You can see what a cell can do, but not what cells actually do," says Cédric Blanpain of the Free University of Brussels, who co-led the skin study[1].

➲ NATURE.COM
For a web focus on cancer metabolism, see:
go.nature.com/hlahwj

All three research groups tried to address this knowledge gap by using genetic techniques to track cells. Parada and his co-workers began by testing whether a genetic marker that labels healthy adult neural stem cells but not their more specialized descendents might also label cancer stem cells in glioblastoma, a type of brain cancer. When they did so, they found that all tumours contained at least a few labelled cells — presumably stem cells. Tumours also contained many unlabelled cells[2]. The unlabelled cells could be killed with standard chemotherapy, but the tumours quickly returned. Further experiments showed that the unlabelled cells originated from ▶

▶ labelled predecessors. When chemotherapy was paired with a genetic trick to suppress the labelled cells, Parada says, the tumours shrank back into "residual vestiges" that did not resemble glioblastoma.

Meanwhile, Hans Clevers, a stem-cell biologist at the Hubrecht Institute in Utrecht, the Netherlands, and his colleagues focused on the gut. They had previously shown that a genetic marker that labels healthy gut stem cells also labels stem cells in benign intestinal tumours, which are precursors of cancer[4]. In their latest study[3], he and his team engineered mice to carry a gene for a drug-inducible marker that, when activated, causes labelled cells to make molecules that fluoresce one of four colours. This experiment yielded single-colour tumours consisting of several cell types, suggesting that each tumour arose from a single stem cell. To check that stem cells continued to fuel the tumours, Clevers added a second, low dose of the drug, triggering a few of the stem cells to change colour. This produced streams of cells in the new colour, showing that stem cells were consistently producing the other cell types.

For the skin study, Blanpain and his group labelled individual tumour cells, without targeting stem cells specifically[1]. They found that cells showed two distinct patterns of division: they either produced a handful of cells before petering out, or went on to produce many cells. Once again, the results pointed to a distinct subset of cells as the engine of tumour growth. What's more, as tumours became more aggressive, they were more likely to produce new stem cells — which can divide indefinitely — and less likely to produce differentiated cells, which can divide only a limited number of times. That could be a key to halting tumour development early, says Blanpain. Rather than eradicating cancer stem cells, for example, therapies could try to coax them to differentiate into non-dividing cells.

The papers provide clear experimental evidence that cancer stem cells exist, says Robert Weinberg, a cancer researcher at the Whitehead Institute in Cambridge, Massachusetts. "They have made a major contribution to validating the concept of cancer stem cells," he says. But cancer cells probably also act in more complex ways than those observed, he warns. For example, non-stem cells within the tumour might de-differentiate into stem cells.

The next step, the three groups say, is figuring out how the cells tracked in these experiments relate to putative cancer stem cells identified by years of transplantation studies. Researchers are already busy hunting for ways to kill these cells; now they have more tools to tell whether such a strategy will work. ■

1. Driessens, G., Beck, B., Caauwe, A., Simons, B. D. & Blanpain, C. *Nature* http://dx.doi.org/10.1038/nature11344 (2012).
2. Chen, J. *et al. Nature* http://dx.doi.org/10.1038/nature11287 (2012).
3. Schepers, A. G. *Science* http://dx.doi.org/10.1126/science.1224676 (2012).
4. Barker, N. *et al. Nature* **457**, 608–611 (2009).

THERAPEUTICS

FDA's claims over stem cells upheld

Drug watchdog wins right to regulate controversial therapies.

BY DAVID CYRANOSKI

A court decision on 23 July could help to tame the largely unregulated field of adult stem-cell treatments. The US District Court in Washington DC affirmed the right of the Food and Drug Administration (FDA) to regulate therapies made from a patient's own processed stem cells. The case hinged on whether the court agreed with the FDA that such stem cells are drugs.

The judge concurred, upholding an injunction brought by the FDA against Regenerative Sciences, based in Broomfield, Colorado. Under the treatment sold by the firm, stem cells are isolated from patients' bone marrow, processed, and the resulting cells injected back into the patients to treat joint pain. The FDA calls this procedure the "manufacturing, holding for sale, and distribution of an unapproved biological drug product", and in August 2010, ordered Regenerative Sciences to stop offering the treatment (see *Nature* **466**, 909; 2010).

"Maintaining the FDA's role as watchdog and regulatory authority is imperative."

During investigations leading up to the injunction, the FDA also found that, because of flaws in its cell processing, the company was violating regulations on "adulteration" that are meant to ensure patients' safety.

Jeanne Loring, a regenerative-medicine scientist at the Scripps Research Institute in La Jolla, California, says that the decision will send a warning to other entrepreneurs offering unapproved stem-cell treatments. "So many people want to start these companies. They say, 'FDA? What FDA?'"

Chris Centeno, the medical director of Regenerative Sciences and one of two majority shareholders, told *Nature* that he plans to appeal against the ruling. During the case, the company claimed that the cells in its 'Regenexx' procedure are not significantly modified before they are reinjected, so the procedure should be considered routine medical practice. The company also argued that because all the processing work is done in Colorado, the procedure should be subject to state law, rather than to regulation by the FDA.

The court disagreed on both counts, noting that "the biological characteristics of the cells change during the process", and that this, together with other factors, means the cells are more than "minimally manipulated".

Leigh Turner, a bioethicist at the University of Minnesota in Minneapolis, agrees. "It is much too simplistic to think that stem cells are removed from the body and then returned to the body without a 'manufacturing process' that includes risk of transmission of communicable diseases," he says. "Maintaining the FDA's role as watchdog and regulatory authority is imperative."

Centeno says that the FDA injunction applies to only one of his company's four stem-cell products — one that requires 4–6 weeks of processing. The procedure will still be available: after the 2010 injunction, the company moved its treatment location to an affiliated Cayman Island clinic.

Centeno plans to continue providing the other three procedures, also used for joint pain, in the United States. In those treatments, the cells are reinjected within two days. Centeno claims that those cells are "minimally manipulated", and that the FDA sees them as the "practice of medicine" and "has no issues" with them. Indeed, until 25 July, a graphic on the Regenerative Sciences website claimed that these three procedures were "FDA approved".

In fact, the FDA has not approved these procedures, and Centeno did not provide documentation to support his claims that the agency views the three treatments as outside its purview. The graphic was removed after *Nature*'s enquiries.

Doug Sipp, a stem-cell ethics and regulation expert at the RIKEN Centre for Developmental Biology in Kobe, Japan, worries that more stem-cell companies might now set up shop outside the United States to avoid regulation, as Regenerative Sciences has done. "Other US stem-cell outfits have close ties with partner clinics in Mexico and other neighbouring countries, which are traditionally regulatory havens for other forms of fringe medicine as well. I suppose it will be business as usual in such places," Sipp says. ■

FAO/T. PALADINI

Mouth lesions are a sign of rinderpest, which has long decimated cattle throughout the world.

INFECTIOUS DISEASE

Officials act to secure cattle-plague virus

Risk of accidental reintroduction shadows rinderpest eradication effort.

BY DECLAN BUTLER

Rinderpest, a devastating cattle disease, has not been seen in the wild for a decade, but it lives on in scores of labs. Twelve months after the world celebrated the success of a years-long vaccination campaign that made rinderpest only the second disease after smallpox to be eradicated, animal-health authorities are turning to the next task: making sure that a lab release — accidental or intentional — doesn't lead to a resurgence.

Rinderpest is as deadly to cattle as highly pathogenic H5N1 avian flu is to chickens. In past decades, outbreaks ripped through herds and wiped out up to 90% of animals, often leaving famine, and sometimes war, in their wake. "Its eradication is a huge, huge, achievement that has happened largely under the radar of most of the virology and scientific community," says David Ulaeto, a member of a seven-person multidisciplinary Joint Advisory Committee (JAC) on rinderpest that was set up to consolidate the eradication by the Rome-based Food and Agriculture Organization of the United Nations (FAO) and the Paris-based World Organisation for Animal Health (OIE).

This October, the JAC will probably issue the first of a series of guidelines for an international oversight system. With the help of ad hoc expert groups, the JAC would approve official repositories of the virus and ensure that they meet tough biosafety standards. The committee would also approve all future research on live rinderpest virus to ensure that its benefits outweigh the risks.

The FAO and OIE don't have the authority to impose such measures on member states, but last year, countries gave the organizations a mandate by endorsing a moratorium on research and declaring that the remaining virus samples should be destroyed or shipped to approved high-security labs. The approach is modelled on the post-eradication phase of the smallpox campaign, which saw the number of labs holding the virus reduced from 76 in 1976 to just 2 in 1984.

To identify labs that might still hold rinderpest virus, the FAO carried out extensive literature searches, liaised with ministries of agriculture and veterinary services worldwide, and wrote "to virtually everyone they could think of", says Ulaeto. By last week, the FAO and OIE had identified some 40 labs. "They were a bit surprised at how many laboratories did have virus," he says.

The list remains confidential, but it includes labs from some 20 countries, thought to be mainly in Africa, the Middle East and Asia, where rinderpest outbreaks were common until recently, and a handful of established rinderpest research centres, such as the Institute for Animal Health in Pirbright, UK, and the Plum Island Animal Disease Center in New York state. One worrying aspect was that some virus samples were found to be held in facilities that had inadequate biosafety levels.

Fears of an accidental release are grounded in experience. After smallpox was eradicated, a lab accident in Birmingham, UK, resulted in two infections and one death. And an accidental release of foot-and-mouth virus from the Pirbright facility, which houses a high-biosecurity, world-reference laboratory for both foot-and-mouth and rinderpest, caused an outbreak in the United Kingdom in 2007.

Active research on rinderpest has waned as the disease has been brought under control over the past few decades, says Michael Baron, a rinderpest researcher at the Pirbright centre. He and others say that the biggest threat is from long-forgotten samples of virus from past research programmes, and serum and other samples collected for diagnostic or other purposes, that may be lurking in lab freezers. Rinderpest vaccine strains, which are stocked in many countries and consist of live attenuated virus, are also a concern. In theory, they could revert to wild type and cause disease outbreaks.

Until the world is certain that rinderpest is gone for good, vaccine strains will probably need to be maintained in high-security labs in several regions so that they can be shipped swiftly to any outbreak, says Baron. But he says that just a couple of pure-research labs would be enough to pursue the valuable scientific opportunities that rinderpest still offers.

Although the virus is closely similar to the human measles virus, for example, cattle don't catch measles and humans don't catch rinderpest. Understanding why this is so could provide insight into the pathology and basic biology of viruses, Baron says. Of more immediate interest, investigators would also like to know whether vaccines can be developed against another related virus, the sheep and goat disease called peste des petits ruminants, that might also protect against rinderpest. That would eliminate the need to keep any stocks of live attenuated rinderpest virus at all.

Baron's home lab contains more than 100 different rinderpest virus isolates, which he says represent "basically the history of the disease". He intends to sequence them all in the next few years — so that they can be recreated if ever needed — and then destroy them. ■

➲ NATURE.COM
To follow the last days of rinderpest, visit:
go.nature.com/1kvuni

PLANETARY SCIENCE

7 MINUTES OF TERROR

The Curiosity rover prepares to plunge down to Mars.

BY ERIC HAND

After an eight-month journey to Mars, success for NASA's Curiosity rover will hinge on a few crucial moments. The largest and most complicated piece of machinery ever sent to the red planet, Curiosity will begin its seven-minute fall through the wispy atmosphere at 05:24 UTC on 6 August. On Earth, mission scientists will be unable to do anything but wait and hope for the signal that the six-wheeled remote laboratory is resting safely in the feeble Martian sunlight.

If Curiosity lands successfully in Gale Crater, it will eventually trundle over to a 5.5-kilometre-tall stack of layered deposits ringed by water-altered minerals. Ascending the mound, the rover will chart hundreds of millions of years of geology and help researchers to deduce whether life could ever have existed on Mars.

But first it has to arrive. On its way down, the spacecraft will fire 76 charges, adopt 6 configurations and slow from 6 kilometres per second to a standstill. It will be the first craft since the Apollo Moon programme of the 1960s and 1970s to use a guided-entry system, and the final leg of the descent will mark the first use of a 'sky crane'. At 900 kilograms, Curiosity is too heavy to land in airbags like earlier rovers, and retrorockets like those used in the Viking Mars landings of the 1970s would kick up damaging dust. Instead, a hovering platform will unspool the rover. "All sorts of things can go wrong," said NASA administrator Charles Bolden at a meeting of the NASA Advisory Council on 25 July. "That's what makes it a real nail-biter."

NASA officials told the Jet Propulsion Laboratory (JPL) in Pasadena, California, to ensure a 95% chance of landing success. Engineers say they have surpassed that: the current assessment, based on millions of simulations, finds only a 1.7% risk of failure. But that holds only if the models have assessed every possible vagary of environment and machine. What's worrisome are the unknown unknowns, says Steven Lee, the mission's guidance, navigation and control-systems manager at the JPL. "Probably the overall biggest risk is our lack of imagination." ■

T–6 min 43 s
GUIDED ENTRY

125 km

5,900 m s^{-1}

Just before entering the atmosphere, the spacecraft sheds two 75-kilogram tungsten weights, shifting its centre of mass and creating a crude ability to generate lift. "We're flying a brick," says Allen Chen, the JPL's operations lead for entry, descent and landing. The ability to change the lift vector lets the spacecraft compensate for unexpected fluctuations in atmospheric density, reducing the size of the potential landing area and making it possible to land in Gale Crater. All the manoeuvres have to be done by dead reckoning, using the spacecraft's internal gyroscopes.

T–5 min 28 s
PEAK HEATING

Lightweight carbon tiles, similar to those on NASA's 1999 Stardust comet-sample-return mission, protect the spacecraft from steel-melting temperatures of up to 2,100 °C.

T–5 min 18 s
PEAK DECELERATION

All parts of the spacecraft had to be designed to withstand accelerations 15 times as strong as Earth's gravity.

Time to touchdown

Altitude

Velocity

T–2 min 28 s
PARACHUTE DEPLOYMENT

11 km

405 m s^{-1}

The parachute, nearly 16 metres in diameter, is expelled by a mortar blast.

Of the 1.7% overall risk of failure, about 1% is due to potential problems with the parachute, such as tangled lines or turbulent oscillations. The parachute design strays little from that used in the successful Viking landings, but there have been only limited tests at the speeds and densities that the parachute will experience in the Martian atmosphere — just a few drops from balloons in Earth's stratosphere during the 1970s.

ROVER EVOLUTION

The size of a small car at almost one tonne, Curiosity is 5 times heavier than the Spirit and Opportunity rovers and 56 times heavier than Sojourner. Unlike its solar-panelled predecessors, it is powered by a radioisotope thermoelectric generator.

Curiosity (2012)

Spirit/ Opportunity (2004)

Sojourner (1997)

T–2 min 4 s
HEAT-SHIELD SEPARATION

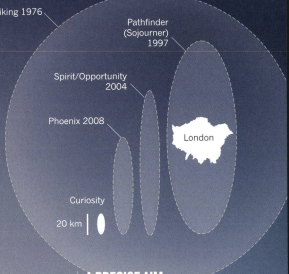
8 km
125 m s^{-1}

After the heat shield is expelled, six radar antennas turn on. They have a field of view of 3 degrees each, and provide the first independent check on the internal estimate of the spacecraft's altitude. "We finally get eyes, and it's absolutely vital," says Steven Lee, Curiosity's guidance, navigation and control-systems manager. To test the system, engineers strapped the antennas to an FA-18 fighter jet and flew it at the ground.

T–53 s
BACK-SHELL SEPARATION

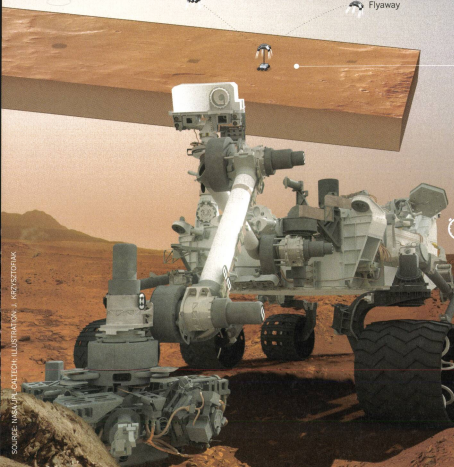
1.6 km
80 m s^{-1}

The parachute is jettisoned, along with the back shell, and eight retrorockets — similar to those used by Viking — begin to fire. They first jerk the spacecraft sideways by 300 metres to get it out of the way of the parachute. "If we didn't do that we'd get hit in the back of the head," says Steve Sell, deputy operations lead for entry, descent and landing at the JPL.



A PRECISE AIM

Thanks to the guided-entry system, the elliptical area in which Curiosity is projected to set down is orders of magnitude smaller than those for previous Mars landers.



T–13 s
SKY CRANE

20 m
0.75 m s^{-1}

Previous landings have used airbags to cushion the final part of the fall. But Curiosity is too heavy, so engineers arranged to use the rover's wheel-suspension system as landing gear. A bridle made of three nylon cords and an 'umbilical cord' for data unspools, dropping the rover 7.5 metres beneath the descent stage, which is still descending under the power of four retrorockets.

T–0:00
TOUCHDOWN

Once the rover stops moving, the bridle cords are severed and the descent stage flies away to land at least 150 metres from the rover. The final touchdown is the second-riskiest part of the landing, after the parachute deployment. The 0.7% risk is divided between different terrain hazards. Rocks and slopes could in rare cases flip the rover over, or Curiosity could land in a crater too deep to escape or on a mesa too steep to descend.

SOURCE: NASA/JPL-CALTECH; ILLUSTRATION: J. KEZWICZYK

MORE ONLINE
NATURE.COM/CURIOSITY

NUTRITIONAL SUPPLEMENTS

Lawsuit challenges anti-ageing claims

Former executive sues manufacturer of pill meant to rejuvenate cells.

BY BRENDAN BORRELL

"You're going to be hearing from my attorney," Brian Egan told his boss on his last day of work, nearly a year ago. Last week, Egan filed a class-action lawsuit that accuses Telomerase Activation Sciences (TA Sciences) in New York of engaging in deceptive business practices in promoting a proprietary herbal extract intended to reverse the effects of ageing.

The lawsuit threatens to put the science of telomeres — repetitive nucleotide sequences that protect the ends of chromosomes during DNA replication — on trial. But Noel Patton, president of TA Sciences, denies all of the allegations. "We stand by what we say," he adds.

The connection between cellular ageing and telomere length is rooted in solid research. Telomeres become shorter every time a cell divides, and when they are lost cells can no longer reproduce. The enzyme telomerase can lengthen telomeres, possibly slowing or reversing degenerative diseases. In one study[1], mice genetically engineered to lack functional telomerase showed brain degeneration and shrunken testes, but those effects were reversed when the enzyme was reactivated.

Such findings have sparked a lot of hype and encouraged a cottage industry of companies that assess a person's 'biological age' on the basis of their telomere length. But TA Sciences has taken the buzz further: it sells a pill called TA-65, which it says can lengthen short telomeres. The pill brings in an annual revenue of US$6 million in the United States alone.

"A compound that can lengthen telomeres would be excellent," says Carol Greider, a molecular biologist at Johns Hopkins University in Baltimore, Maryland, who shared a Nobel prize for her work on how telomerase protects chromosomes. But, she adds, "we would need to test it rigorously".

The active ingredient of TA-65 was isolated from the herb *Astragalus membranaceus* and patented by Geron, a biopharmaceutical firm in Menlo Park, California. Research sponsored by TA Sciences and other companies has shown that the compound can lengthen telomeres in mice[2] and humans[3], but Greider and others are sceptical of the assay used.

Calvin Harley, president of Telome Health in Menlo Park, spearheaded the studies as chief scientific officer at Geron. He stands by the conclusion that TA-65 is a "weak telomerase activator". However, TA Sciences sells the pill as a nutritional supplement, or 'nutraceutical', rather than a drug, so the firm's health claims have not been evaluated by the US Food and Drug Administration (FDA).

DIFFICULT RELATIONSHIP

In May 2011, Patton hired Egan to help to expand TA Sciences' reach in foreign markets. Egan was required to take TA-65 twice a day, he later wrote in a discrimination complaint, so "that I could tell customers that I was also taking the product, and that it was safe and effective". Patton denies that it was obligatory.

On 14 September, Egan says, he told Patton that he had been diagnosed with prostate cancer. The next day, according to Egan, Patton fired him and said that his prostate cancer could ruin the company. Egan says that when he was fired, he was offered a cash settlement to keep quiet about his cancer, but turned it down.

Patton denies Egan's version of events. According to an affidavit that Patton filed in the discrimination suit, Egan was fired for meagre sales. On being told of his dismissal, Patton alleges in the affidavit, Egan threw his keys at his boss and demanded to settle things "man to man". An employee broke up the confrontation, and Egan stormed out, says Patton. Egan denies Patton's version of events.

On 19 September, Egan told a potential TA Sciences partner in Spain that he had developed cancer while taking TA-65. Patton and TA Sciences sued Egan for defamation in March, saying that he had lost the company $2 million in sales. Patton says that TA Sciences believes that if Egan had cancer, he had it before he started taking TA-65.

"A compound that can lengthen telomeres would be excellent. But we would need to test it rigorously."

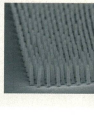
Egan stands by his allegations, and has now launched a broader attack on the company's science with his class-action suit, which he filed on 23 July, in the New York State supreme court, along with another man who took TA-65. The suit challenges statements on TA Sciences' website, including the assertion that TA-65 can lengthen short telomeres.

Greider doubts that TA-65 caused Egan's cancer, but agrees that the science behind it is murky. A telomere-lengthening compound would be a boon to patients dying of bone-marrow failure and pulmonary fibrosis, she says, and firms could be expected to explore its pharmaceutical potential. "I don't think a company would be selling it on the side as a nutraceutical," she says. ■

1. Jaskelioff, M. *et al. Nature* **469,** 102–106 (2011).
2. Bernardes de Jesus, B. *et al. Aging Cell* **10,** 604–621 (2011).
3. Harley, C. B. *et al. Rejuvenation Res.* **14,** 45–56 (2011).

MORE ONLINE

TOP STORY



Electronic sensor rivals sensitivity of human skin go.nature.com/lsxuqc

MORE NEWS

- Dormant HIV gets rude awakening go.nature.com/yvssay
- Storms may speed ozone loss above the United States go.nature.com/qwmcf4
- The skin's secret surveillance system go.nature.com/slhnr2

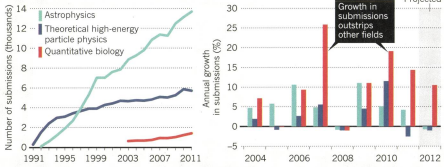
VIDEO

Termites explode to defend their colony go.nature.com/i2vxrp

R. HANUS

BIOLOGY OPENS UP

Quantitative-biology papers are a small but rising fraction of submissions to the ArXiv preprint database.



SOURCE: ARXIV.ORG

PUBLISHING

Geneticists eye the potential of arXiv

Population biologists turn to pre-publication server to gain wider readership and rapid review of results.

BY EWEN CALLAWAY

The preprint server arXiv.org is perhaps best known as the preserve of theoretical physicists and astrophysicists. But 2008 saw an influx of submissions of unpublished manuscripts, or preprints, by condensed-matter physicists who wanted to stake claims to the fast-moving subject of iron-based superconductors called pnictides. Now the life sciences may be on the cusp of their own 'pnictide moment', with population geneticists leading the charge.

In the past month, leading research groups have posted to arXiv high-profile papers on the genetic history of southern Africans[1] and Europeans[2]. Other prominent population geneticists have submitted methods-based papers to the server, which is hosted by Cornell University in Ithaca, New York. The number of biology papers on the server is still small in comparison with physical-sciences preprints (see 'Biology opens up'), but Paul Ginsparg, a theoretical physicist at Cornell who founded arXiv in 1991 (ref. 3), welcomes what he hopes could be a sea change.

"It's wonderful if biologists are belatedly joining the late twentieth century," he quips. "Welcome to the party; better late than never."

Life-sciences papers have existed on arXiv almost since its inception, but biologists have typically shied away from this approach amid fears of getting scooped, or of offending journals. It was only in 2003 that the site inaugurated a section specifically for papers on quantitative biology, or 'q-bio' for short. Yet papers posted to the section in the past tended to report esoteric models and methods, often from physical scientists dabbling in biology.

Rich in mathematics and steeped in the open-data traditions of genomics, population genetics would seem the ideal candidate to dip its toes into pre-publication, which brings the advantages of speed and open discussion. "I grew up in the physics community," says Richard Neher, a population geneticist at the Max Planck Institute for Developmental Biology in Tübingen, Germany, "and putting things up on arXiv is a natural thing for me to do." Neher has co-authored more than ten submissions to arXiv since 2004.

Despite such examples, one of the best-known life-sciences preprints on arXiv comes from microbiology, not population genetics. Posted as a rebuttal to a 2011 *Science* paper reporting that a strain of *Halomonas* bacteria from a Californian lake could incorporate arsenic into its DNA[4], the preprint appeared on the server in January[5] before its publication in *Science* this month[6]. Yet a population geneticist still played a part — the paper's co-author, Leonid Kruglyak, of Princeton University in New Jersey and the Howard Hughes Medical Institute, works in the field. Kruglyak says he will consider arXiv for future studies from his lab.

Another attention-grabbing submission by prominent geneticists, posted on 23 July, compares genomic variation in 22 African populations to suggest an ancient genetic link between people in southern and eastern Africa[1]. One of the paper's senior authors, geneticist David Reich of Harvard Medical School in Boston, Massachusetts, publishes routinely in *Nature* and the *Public Library of Science* journals, and co-author Carlos Bustamante, of Stanford University School of Medicine in California, is a leader in the field. Reich says that first author Joseph Pickrell, also at Harvard Medical School, suggested using arXiv. Reich and the other co-authors saw no good reason not to post the manuscript there. "It could be an example of the younger generation coming in and finding this sort of thing natural," says Ginsparg.

For most life scientists, however, pre-publication is still "more of a trickle" than a trend, says statistical geneticist Graham Coop of the University of California, Davis. He and his postdoc, statistician Peter Ralph, posted a paper on 16 July analysing genetic relatedness among neighbouring European populations, and Coop remains bullish about arXiv's potential. "Biology will soon have to embrace this trend fully: the speed of discussion, comment and pre-publication review allowed is needed in biology more than most fields," he says.

Yet some population geneticists still feel that those posting to arXiv are sticking their necks out. Some biology journals, such as those published by the Ecological Society of America in Washington DC, for example, expressly prohibit pre-publication in citable public archives. And there are the concerns over establishing who was first to a discovery.

But Ginsparg says that pre-publication is more likely to stop scientists from being scooped. In many physics fields, publication on arXiv is what counts for claiming priority, and journal reviewers can use the server to check that discoveries are correctly attributed. An authoring history that accompanies all arXiv papers also allows scientists to arbitrate disputes over priority. In the 21 years since arXiv began, Ginsparg has seen astrophysicists, computer scientists and others go from sceptics to devotees. "Once a community adopts arXiv, it never seems to relinquish it," he says. ■

1. Pickrell, J. K. *et al.* Preprint at http://arxiv.org/abs/1207.5552 (2012).
2. Ralph, P. & Coop, G. Preprint at http://arxiv.org/abs/1207.3815 (2012).
3. Ginsparg, P. *Nature* **476,** 145–147 (2011).
4. Wolfe-Simon, F. *et al. Science* **332,** 1163–1166 (2011).
5. Reaves, M. L. *et al.* Preprint at http://arxiv.org/abs/1201.6643 (2012).
6. Reaves, M. L. *et al. Science* **337,** 470–473 (2012).

CORRECTIONS

The graph in the News story 'Gene data to hit milestone' (*Nature* **487,** 282–283; 2012) miscounted data sets in ArrayExpress for the years 2003–11. The corrected graph can be seen online at go.nature.com/2wrlpx.

The News Feature 'Beta test' (*Nature* **487,** 160–162; 2012) gave the wrong affiliation for Stefan Schönert. He is at the Technical University Munich.

Physical scientists protested against funding reforms with a mock funeral for British science.

Duel to the death

Physicists, chemists and mathematicians in the United Kingdom are furious about funding reforms that they say threaten blue-skies research.

BY ANANYO BHATTACHARYA

The horse-drawn Victorian hearse canters past London's Houses of Parliament and round Parliament Square before trotting smartly down Whitehall. Straggling behind it comes an eclectic mix of scientists, ranging from students with body piercings to tweed-jacketed professors. The spectacle makes an odd addition to the spring afternoon traffic, bringing to mind an elegant, if unusual, state funeral.

The truth, however, is much stranger. "We're protesting," explains one PhD student to a puzzled tourist, "against our research funder."

That funder is the Engineering and Physical Sciences Research Council (EPSRC), the government body that holds the biggest public purse for physics, mathematics and engineering research in the United Kingdom. Facing a growing cash squeeze and pressure from the government to demonstrate the economic benefits of research, in 2009 the council's chief executive, David Delpy, embarked on a series of controversial reforms.

Some were intended to cut the overwhelming number of grant proposals that the EPSRC receives, by limiting resubmissions and temporarily blocking people who submitted too many unsuccessful applications from sending in more. A second set of reforms included requirements that grant applicants explain how their research might generate economic or other benefits, as well as a vigorous overhaul of the EPSRC's research portfolio.

The changes incensed many physical scientists, who protested that the policy to blacklist grant applicants was draconian. They complained that the EPSRC's decision to exert more control over the fields it funds risked sidelining peer review and would favour short-term, applied research over curiosity-driven, blue-skies work in a way that would be detrimental to British science. The souring relationship between the EPSRC and parts of its constituency reached a conspicuously public nadir in May, when

DAN KITWOOD/GETTY IMAGES

disaffected researchers launched the 'Science for the Future' campaign with the hearse stunt, which ended by delivering the coffin, signifying the death of British science, and a petition demanding the "immediate reform of the EPSRC's policies" to the prime minister in Downing Street. In a letter to *The Daily Telegraph* newspaper in support of the protestors, nine Nobel laureates in the United Kingdom and United States accused the EPSRC of "manipulating the process of peer review" and "establishing favouritism schemes".

The battle could be played out elsewhere. Many government funding bodies are facing diminishing budgets in the wake of the global financial crisis, and increasing pressure from politicians to show that the research they are funding will contribute to economic growth. "This is a challenge for research funding agencies across the globe," says Julia Lane, an expert in science policy who formerly worked at the National Science Foundation (NSF) in Arlington, Virginia, and is now a senior managing economist at the American Institutes for Research in Washington DC. "And they're all struggling to provide enough information for policy-makers so that they can keep funding going for basic research."

In Britain, it looks likely that researchers will have to live with this new reality: Delpy and his supporters say that the policies are an unavoidable reaction to a slumping budget and that there will be no U-turns. According to Delpy's detractors, however, the EPSRC provides a lesson in how not to implement such reforms. "They shot themselves in the foot by alienating the community that they're here to serve," says Paul Clarke, a synthetic organic chemist at the University of York, and one of the EPSRC's most vociferous critics.

OPENING ROUND

The tensions started rising in 2007, shortly after Delpy, a physicist by training, left his post managing the research portfolio at University College London to take command of the EPSRC and its research budget of some £800 million (US$1.3 billion). Delpy faced a problem: an overwhelming number of grant applications and a flat budget were starting to push up rejections. In 2008, the success rate for applications dropped from a typical 30% to 26% overall and even lower in some fields (see 'Physics of funding'). In March 2009, the EPSRC announced that from the following month researchers could no longer resubmit grant proposals that had been rejected, a policy that has since been implemented at several other UK research councils. At around the same time, it also barred researchers with a record of rejections from sending in any further funding applications for 12 months — a policy that initially hit more than 200 people.

The changes outraged physical scientists. Within 2 weeks of the blacklisting policy being announced, more than 1,200 people had signed an online petition demanding that it be rescinded (see *Nature* **458,** 391; 2009). In May 2009, the EPSRC was forced to water down the policy by delaying its introduction to April 2010, and allowing researchers to apply for one grant during the 12-month 'cooling-off' period. The EPSRC says that only ten researchers are currently blacklisted.

Clarke is one of those ten. Six years ago he concurrently held three EPSRC grants — more than any other researcher at the same career stage. But in January, after submitting a third unsuccessful proposal, he was told that he had been automatically barred from sending in more. Clarke, who works on the chemical origins of life and on synthesizing natural compounds, says that before the EPSRC banned resubmissions, he was able to address critiques of rejected proposals and send them back for re-review — after which they would often be successful. Rather than being a problem, he views the large number of proposals as a sign of scientists with many good ideas, and advocates widening the pool of reviewers by mandating that those who submit EPSRC grants also review a certain number of proposals each year.

The EPSRC was also under growing pressure to demonstrate the impact of its investment in research. This grew out of the 2006 Warry report, *Increasing the Economic Impact of Research Councils*, as well as subsequent reports, demanding that all the UK research councils show that they are getting the most bang for their buck. In 2009, the research councils started to require that funding applicants submit a two-page Pathways to Impact statement, summarizing how they intended to maximize the societal or economic benefits of a project — through commercialization of results, for example, or through public outreach. In November 2011, the EPSRC added a 'national importance' criterion, requiring that researchers describe in a separate statement "the extent to which the research proposed has the potential, over 10–50 years, to meet national strategic needs".

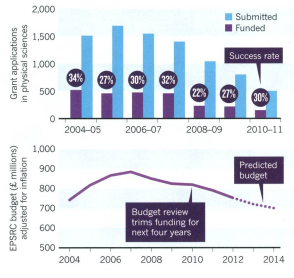
Similar requirements are already routine at some other funding agencies, such as the NSF, which has asked applicants to explain their proposed project's 'broader impacts' on science and society since 1997. That requirement has come in for some criticism (see *Nature* **475,** 141; 2011) — but not the level of anger that the new request seemed to inspire in Britain's physical scientists. The impact statements, they said, showed that the EPSRC is inappropriately favouring short-term projects that will have economic benefits. "It is changing the fundamental ethos of research to make it more responsive to the market," says physicist Philip Moriarty of the University of Nottingham, an outspoken critic of the EPSRC reforms. And the national-importance criterion, say some, is simply asking for the impossible. "It's very hard to justify the economic importance of work that might not become applicable to real-world problems for decades," says postdoctoral mathematician Will Merry at the University of Cambridge, UK. (Merry's research problem, how to solve the motions of three or more interacting bodies, was first set out by Isaac Newton in 1687 — suggesting that science sometimes has to take the long road.)

Martin Rees, former president of the Royal Society in London and a cosmologist at the University of Cambridge, says he worries that the new requirements could affect how young researchers apply for their first grant. "It's going to make them slant their application in a way that might not be optimal from the point of view of the research," he says. The council should focus on making sure that the "brightest people don't get discouraged", adds Rees, who says he finds the idea of asking researchers to write about the potential future national importance of their work "absurd".

But the EPSRC had more reforms up its sleeve. In July 2011, it published the first of three phases in its Shaping Capability strategy, which divided the organization's research portfolio into more than 100 fields, with the aim of maintaining or expanding areas of national importance and excellence, such as catalysis and energy storage, and shrinking others such

PHYSICS OF FUNDING

The UK's main physical-science funder introduced policies to cut grant submissions in 2009 (top), partly in response to a declining budget (bottom).



SOURCE: TOP: EPSRC/C. HAYES/UNIV. NOTTINGHAM; BOTTOM: EPSRC

as mathematical physics and mobile computing. The council revealed that future postdocs and other fellowships would be funded only in areas that were in line with the strategy; at the outset, for example, mathematics postdocs would be funded only in statistics and applied probability.

To the critics, that decision was further evidence that the council ranked short-term pay-off above blue-skies research, in this case favouring the needs of the City, London's large financial sector, which is hungry for statistical expertise. The policy had an immediate impact for Merry, who finished a 12-month EPSRC doctoral-prize fellowship and struggled to find a postdoc at home. He is now set to start one at ETH Zurich in September. Merry says that all the maths PhD students he knows at the University of Cambridge are following a similar path. "The most visible outcome of the change in funding is that to the best of my knowledge they're all going abroad next year," he says.

Some researchers also have a broader concern: that EPSRC administrators are taking over the role of working scientists in deciding how money should be spent and, as a consequence, are funding mediocre research in arbitrarily selected areas. Instead, they say, the EPSRC should focus on supporting the best-quality science in any area of its remit, as judged by peer review. "Non-scientists are making decisions that impact on the future spend of science money — and that is wrong," says Tony Barrett, a synthetic organic chemist at Imperial College London and one of the organizers of the Science for the Future campaign. "They are not qualified to make those decisions."

The uproar over impact and national importance grew so loud that, in November 2011, the House of Lords Science and Technology Committee held a session to discuss the policies with Delpy and civil engineer John Armitt, then the council's chair. The EPSRC also agreed to further discussions with the scientific community before rolling out the final phase of the Shaping Capability strategy, which it did in March 2012. The announcement, which filled in funding details for just over 50 fields, met a frosty, but somewhat less angry, reception.

MAKING IMPACT

Delpy stoutly defends his organization and its reforms. The policies to cut grant applications have worked, he says: the success rates for applications are back up to around 30–35% — "a healthy level of competition".

He rejects the criticisms of the impact and national-importance strategies. The focus on economic impacts is nothing new, he says, nor does it come from his own experience in applied bioengineering. (He developed techniques for monitoring premature babies.) The 1994 royal charter that officially established the EPSRC states that contributing to the country's economic competitiveness is one of the body's three main aims. The prominence placed on it in recent years is largely the result of government pressure, Delpy says. "You must realize that the term 'economic impact' was something that was imposed on us by the Treasury." What's more, the EPSRC does not expect a precise forecast of what impacts a research proposal could have decades down the line, he says; rather, researchers are encouraged to describe how any impacts that they can foresee might be speedily achieved.

SANG TAN/AP

In May, physical scientists called for research-council reform at the House of Commons in London.

Delpy maintains that peer review is paramount at the agency. The decision to limit the EPSRC's maths postdocs to statistics was based on an independent international review of maths commissioned by the council and published in 2010, which singled out the field as an area of serious concern. "We decided we had to do something to get new blood into statistics," he says. "This is not a response to government or the call of the City." Mathematicians can turn to the Royal Society or the Leverhulme Trust in London for funding, he points out. And although the research council's budget is set to decline — by 6% in cash terms between 2010–11 and 2014–15 — he says that the proportion of funding going to discovery-led or blue-skies research has stayed roughly flat at around 50–60% of the total. "Where's the evidence the system is broken?" he asks.

Delpy does regret the rift that the reforms have opened up between the council and researchers. "Of course we could have done things better," he concedes, particularly in communicating what the council was doing and why. "I would have liked to have been able to carry the community with me and have everyone feel that they were engaged." At the same time, he says, "if you change any of the ways of working you are going to create some degree of upset".

Britt Holbrook, a philosopher at the University of North Texas in Denton, who specializes in science policy, says that the EPSRC's biggest mistake lay in banning resubmissions and blacklisting researchers. "It's totally understandable that did not go down well," he says. "If you stop people from resubmitting, that cuts out a large part of the value of peer review." But Holbrook, who co-authored an influential review of the NSF's broader-impacts criterion, says that the critics are fighting a losing battle against the impact agenda. "People are working against the political and economic realities," he warns, and would be better off trying to shape the drive for impact, rather than block it.

The hearse parked outside Downing Street, however, suggests that this advice is falling on deaf ears; the organizers of Science for the Future say that they are planning further protests over the summer. And Delpy's critics have little time for his explanations, saying, for example, that they are sceptical of EPSRC figures showing no decline in blue-skies research, given that the council defines the term itself. Clarke prefers to point to figures showing that in 2010–11, the EPSRC funded 151 proposals in the physical sciences, excluding engineering, down from more than 500 in 2004–05. Yet there are more than 3,000 scientists in the United Kingdom who are eligible to apply for EPSRC funding.

"People would rather not submit than submit, get blacklisted and hence be seen as a failing academic by their department," says Clarke. "There is now a culture of fear in academic departments." ■

Ananyo Bhattacharya *is* Nature*'s chief online editor.*

nature MIDDLE EAST
Emerging science in the Arab world

twitter facebook google+

The Arab world has a rich history of scientific breakthroughs. Ensure you're part of the scientific and medical community gaining access to a regularly updated list of **science jobs and local events in the Middle East**.

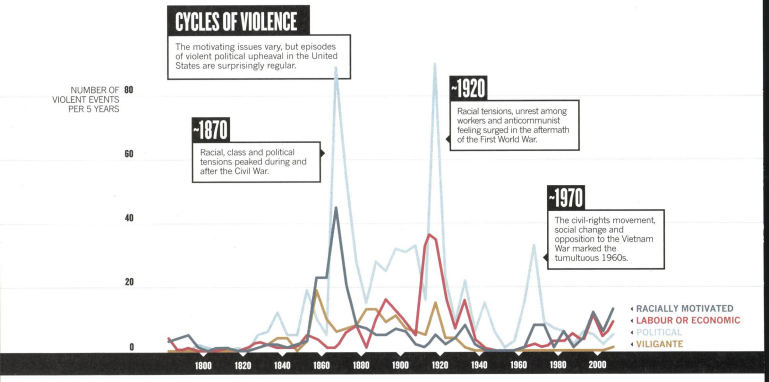
From the latest news, in-depth features, expert commentaries and research highlights to regional jobs, event listings and alerts direct to your inbox – ***Nature Middle East*** is your free comprehensive portal keeping you at the centre of science in the Arabic speaking Middle East.

nature.com/nmiddleeast

Download the free app for your phone at http://gettag.mobi

Sponsored by

nature publishing group npg



HISTORY AS SCIENCE

Advocates of 'cliodynamics' say that they can use scientific methods to illuminate the past. But historians are not so sure.

BY LAURA SPINNEY

Sometimes, history really does seem to repeat itself. After the US Civil War, for example, a wave of urban violence fuelled by ethnic and class resentment swept across the country, peaking in about 1870. Internal strife spiked again in around 1920, when race riots, workers' strikes and a surge of anti-Communist feeling led many people to think that revolution was imminent. And in around 1970, unrest crested once more, with violent student demonstrations, political assassinations, riots and terrorism (see 'Cycles of violence').

To Peter Turchin, who studies population dynamics at the University of Connecticut in Storrs, the appearance of three peaks of political instability at roughly 50-year intervals is not a coincidence. For the past 15 years, Turchin has been taking the mathematical techniques that once allowed him to track predator–prey cycles in forest ecosystems, and applying them to human history. He has analysed historical records on economic activity, demographic trends and outbursts of violence in the United States, and has come to the conclusion that a new wave of internal strife is already on its way[1]. The peak should occur in about 2020, he says, and will probably be at least as high as the one in around 1970. "I hope it won't be as bad as 1870," he adds.

SOURCE: REF. 1

Turchin's approach — which he calls cliodynamics after Clio, the ancient Greek muse of history — is part of a groundswell of efforts to apply scientific methods to history by identifying and modelling the broad social forces that Turchin and his colleagues say shape all human societies. It is an attempt to show that "history is not 'just one damn thing after another'", says Turchin, paraphrasing a saying often attributed to the late British historian Arnold Toynbee.

Cliodynamics is viewed with deep scepticism by most academic historians, who tend to see history as a complex stew of chance, individual foibles and one-of-a-kind situations that no broad-brush 'science of history' will ever capture. "After a century of grand theory, from Marxism and social Darwinism to

structuralism and postmodernism, most historians have abandoned the belief in general laws," said Robert Darnton, a cultural historian at Harvard University in Cambridge, Massachusetts, in a column written in 1999.

Most think that phenomena such as political instability should be understood by constructing detailed narratives of what actually happened — always looking for patterns and regularities, but never forgetting that each outbreak emerged from a particular time and place. "We're doing what can be done, as opposed to aspiring after what can't," says Daniel Szechi, who studies early-modern history at the University of Manchester, UK. "We're just too ignorant" to identify meaningful cycles, he adds.

But Turchin and his allies contend that the time is ripe to revisit general laws, thanks to tools such as nonlinear mathematics, simulations that can model the interactions of thousands or millions of individuals at once, and informatics technologies for gathering and analysing huge databases of historical information. And for some academics, at least, cliodynamics can't come a moment too soon. "Historians need to abandon the habit of thinking that it's enough to informally point to a sample of cases and to claim that observations generalize," says Joseph Bulbulia, who studies the evolution of religion at Victoria University of Wellington in New Zealand.

BETTMANN/CORBIS

TOPICAL PRESS AGENCY/GETTY

N. BOENZI/NEW YORK TIMES CO./GETTY

Periods of rioting and upheaval have recurred roughly every 50 years in US history.

FROM ECOLOGY TO HISTORY

Turchin conceived cliodynamics during what he jokingly calls a midlife crisis: it was 1997, he was 40 years old, and he had come to feel that all the major ecological questions about population dynamics had been answered. History seemed to be the next frontier — perhaps because his father, the Russian computer scientist Valentin Turchin, had also wondered about the existence of general laws governing societies. (The elder Turchin's dissident writings about the origins of totalitarianism were among the reasons that the Soviet Union exiled him in 1977, after which he moved his family to the United States.)

What is new about cliodynamics isn't the search for patterns, Turchin explains. Historians have done valuable work correlating phenomena such as political instability with political, economic and demographic variables. What is different is the scale — Turchin and his colleagues are systematically collecting historical data that span centuries or even millennia — and the mathematical analysis of how the variables interact.

In their analysis of long-term social trends, advocates of cliodynamics focus on four main variables: population numbers, social structure, state strength and political instability. Each variable is measured in several ways. Social structure, for example, relies on factors such as health inequality — measured using proxies including quantitative data on life expectancies — and wealth inequality, measured by the ratio of the largest fortune to the median wage. Choosing appropriate proxies can be a challenge, because relevant data are often hard to find. No proxy is perfect, the researchers concede. But they try to minimize the problem by choosing at least two proxies for each variable.

Then, drawing on all the sources they can find — historical databases, newspaper archives, ethnographic studies — Turchin and his colleagues plot these proxies over time and look for trends, hoping to identify historical patterns and markers of future events. For example, it seems that indicators of corruption increase and political cooperation unravels when a period of instability or violence is imminent. Such analysis also allows the researchers to track the order in which the changes occur, so that they can tease out useful correlations that might lead to cause–effect explanations.

ENDLESS CYCLES

When Turchin refined the concept of cliodynamics with two colleagues — Sergey Nefedov of the Institute of History and Archaeology in Yekaterinburg, Russia, and Andrey Korotayev of the Russian State University for the Humanities in Moscow — the researchers found that two trends dominate the data on political instability. The first, which they call the secular cycle, extends over two to three centuries. It starts with a relatively egalitarian society, in which supply and demand for labour roughly balance out. In time, the population grows, labour supply outstrips demand, elites form and the living standards of the poorest fall. At a certain point, the society becomes top-heavy with elites, who start fighting for power. Political instability ensues and leads to collapse, and the cycle begins again.

Superimposed on that secular trend, the researchers observe a shorter cycle that spans 50 years — roughly two generations. Turchin calls this the fathers-and-sons cycle: the father responds violently to a perceived social injustice; the son lives with the miserable legacy of the resulting conflict and abstains; the third generation begins again. Turchin likens this cycle to a forest fire that ignites and burns out, until a sufficient amount of underbrush accumulates and the cycle recommences.

These two interacting cycles, he says, fit patterns of instability across Europe and Asia from the fifth century BC onwards. Together, they describe the bumpy transition of the Roman Republic to the Roman Empire in the first century BC. He sees the same patterns in ancient Egypt, China and Russia, and says that they explain the timing of last year's Egyptian uprising, which took the regime of then-president

Hosni Mubarak by surprise. At the time, the Egyptian economy was growing and poverty levels were among the lowest in the developing world, so the regime could reasonably have expected stability. In the decade leading up to the revolution, however, the country saw a quadrupling of graduates with no prospects — a marker of elite overproduction and hence, Turchin argues, trouble.

Turchin has also applied this approach to other historical puzzles, such as how religions grow. Several models have been proposed. One is that they grow in a linear fashion as nonbelievers spontaneously 'see the light'. Another model holds that the number of converts increases exponentially, like infections with a contagious disease, as outsiders come into contact with growing numbers of converts. Using several independent proxies, Turchin has mapped conversions to Islam in medieval Iran and Spain, and found that the data fit the contagion model most closely[2]. Using the same techniques, he has also shown that the model describes the expansion of Christianity in the first century AD, and of Mormonism since the Second World War.

Claudio Cioffi-Revilla, a computer social scientist at George Mason University in Fairfax, Virginia, welcomes cliodynamics as a natural complement to his own field: doing simulations using 'agent-based' computer models. Cioffi-Revilla and his team are developing one such model to capture the effects of modern-day climate change on the Rift Valley region in East Africa, a populous area that is in the grip of a drought. The model starts with a series of digital agents representing households and allows them to interact, following rules such as seasonal migration patterns and ethnic alliances. The researchers have already seen labour specialization and vulnerability to drought emerge spontaneously, and they hope eventually to be able to predict flows of refugees and identify potential conflict hotspots. Cioffi-Revilla says that cliodynamics could strengthen the model by providing the agents with rules extracted from historical data.

GLOBAL TRENDS

Cliodynamics has another ally in Jack Goldstone, director of the Center for Global Policy at George Mason University and a member of the Political Instability Task Force, which is funded by the US Central Intelligence Agency to forecast events outside the United States. Goldstone has searched for cliodynamic patterns in past revolutions, and predicts that Egypt will face a few more years of struggle between radicals and moderates and 5–10 years of institution-building before it can regain stability. "It is possible but rare for revolutions to resolve rapidly," he says. "Average time to build a new state is around a dozen years, and many take longer."

But Goldstone cautions that cliodynamics is useful only for looking at broad trends. "For some aspects of history, a scientific or cliodynamic approach is suitable, natural and fruitful," he says. For example, "when we map the frequency versus magnitude of an event — deaths in various battles in a war, casualties in natural disasters, years to rebuild a state — we find that there is a consistent pattern of higher frequencies at low magnitudes, and lower frequencies at high magnitudes, that follows a precise mathematical formula." But when it comes to predicting unique events such as the Industrial Revolution, or the biography of a specific individual such as Benjamin Franklin, he says, the conventional historian's approach of assembling a narrative based on evidence is still best.

Herbert Gintis, a retired economist who is still actively researching the evolution of social complexity at the University of Massachusetts Amherst, also doubts that cliodynamics can predict specific historical events. But he thinks that the patterns and causal connections that it reveals can teach policy-makers valuable lessons about pitfalls to avoid, and actions that might forestall trouble. He offers the analogy of aviation: "You certainly can't predict when a plane is going to crash, but engineers recover the black box. They study it carefully, they find out why the plane crashed, and that's why so many fewer planes crash today than used to."

None of these arguments, however, has done much to soften scepticism among historians in general. The essential weakness of any attempt to make predictions based on trends, says Szechi, is the appalling patchiness of historical information. Records can be preserved or destroyed by chance: in 1922, for example, fighting in the Four Courts area of Dublin during the Irish Civil War led to a fire that destroyed the country's entire medieval archive. More generally, says Szechi, knowledge tends to pool around narrow subject areas. "We can tell you in great detail what the grain prices were in a few towns in southern England in the Middle Ages," he says. "But we can't tell you how most ordinary people lived their lives."

Concerted efforts are now under way to fill those holes. Harvey Whitehouse, an anthropologist at the University of Oxford, UK, is overseeing the construction of a database of information about rituals, social structure and conflict around the globe since records began. It is a huge undertaking, involving historians, archaeologists, religious scholars, social scientists and even neuroscientists, and it will take decades to complete — assuming that funding can be found beyond the UK government's current 5-year commitment. But Whitehouse believes that the research that is feeding the database will complement Turchin's approach by throwing light on the immediate triggers of political violence. He argues[3], for example, that for such violence to happen, individuals must begin to identify strongly with a political group. One powerful way for groups to cement that identification is through rituals, especially frightening, painful or otherwise emotional ones that create a body of vivid, shared memories.

> "MOST HISTORIANS HAVE ABANDONED THE BELIEF IN GENERAL LAWS."

"People form the impression that the most profound insights they have into their own personal history are shared by other people," says Whitehouse, who explored this fusion of identities in an as-yet unpublished survey of revolutionary brigades in Misrata, Libya, last December, along with his colleague Brian McQuinn, an anthropologist at Oxford who studies civil wars. Only once such fusion has occurred do people become willing to fight and die for the group, he says. Therefore, if Turchin's prediction of unrest in the United States around 2020 is correct, Whitehouse would expect the next few years to see an increase in tightly knit US groups whose rituals have a threatening quality but promise great rewards.

Turchin can't say who those groups might be, what cause they will be fighting for or what form the violence will take. Previous bouts of turbulence were not dominated by any one issue, he says. But he already sees the warning signs of social strife, including a surplus of graduates and increasing inequality. "Inequality is almost always a bad thing for societies," he says.

That said, Turchin insists that the violence is no more inevitable than an outbreak of measles. Just as an epidemic can be averted by an effective vaccine, violence can be prevented if society is prepared to learn from history — if the US government creates more jobs for graduates, say, or acts decisively to reduce inequality.

But perhaps revolution is the best, if not the only, remedy for severe social stresses. Gintis points out that he is old enough to have taken part in the most recent period of turbulence in the United States, which helped to secure civil rights for women and black people. Elites have been known to give power back to the majority, he says, but only under duress, to help restore order after a period of turmoil. "I'm not afraid of uprisings," he says. "That's why we are where we are." ■

Laura Spinney *is a freelance writer in Lausanne, Switzerland.*

1. Turchin, P. *J. Peace Res.* **49**, 577–591 (2012).
2. Turchin, P. *Historical Dynamics* (Princeton Univ. Press, 2003).
3. Swann, W. B. Jr, Jetten, J., Gómez, Á., Whitehouse, H. & Bastian, B. *Psychol. Rev.* **119**, 441–456 (2012).

COMMENT

BRAIN Extraordinary tale of Francis Crick teaching Galileo consciousness theory **p.29**

CULTURE Exhibition celebrates African views of the cosmos **p.30**

EVOLUTION A brief look at curious human behaviour, from tickling to burping **p.31**

ECOLOGY Post mortems show bats killed by wind-turbine blades, not air pressure **p.32**

NASA/JPL

SAMPEX, the first of NASA's inexpensive Small Explorer satellites in orbit, is due back to Earth this year.

Let academia lead space science

NASA must put more of its money into thrifty missions led by principal investigators, says **Daniel N. Baker**.

The Mars Curiosity rover, which all space scientists fervently hope will touch down on the red planet safely this week, is a prime example of an expensive and complicated NASA mission. With a landing scheme involving 76 pyrotechnic devices firing on time and a US$2.5-billion price tag, it is a high-risk endeavour. By contrast, the Mars Atmosphere and Volatile Evolution Mission (MAVEN) is a project being run out of our laboratory in Colorado to explore Mars's upper atmosphere and ionosphere. It is set to launch in 2013 for about $500 million. It is on budget, on schedule and promises compelling science. Yet the Scout programme, under which such small Mars missions were funded, has recently been axed.

The planetary exploration flagship programmes and the vastly over-budget James Webb Space Telescope are symptomatic of a core problem in space research. Increasingly, NASA's focus is on big projects that promise to return tremendous science benefits. But these programmes absorb most of the available funding for space research. They shift resources away from efficient and effective principal investigators (PIs) at universities, an approach in which a single person is responsible to NASA for the success of a mission, and towards bureaucratic NASA centres. This is the wrong direction for space research, especially in a time of scarce funding.

In my opinion, we need to turn civilian space-policy thinking on its head. Missions managed by PIs should be the highest priority for NASA, not the lowest. I am not talking about the 'faster, better, cheaper' approach of the 1990s, with skeleton crews of engineers at NASA centres. I am talking about missions led by university scientists with a real passion for research. This strategy would reduce budgetary overruns, increase the frequency of launches and enhance excitement like few other things could.

THREE-WAY PARTNERSHIP

At its beginning, the US civilian space programme was crafted as a three-way partnership between government (NASA), industry and academia. From the famous 1945 report of engineer and policy adviser Vannevar Bush, *Science — The Endless Frontier*, through to the ▶

unflinching commitment of NASA's second administrator, James Webb, the founding fathers of space research put universities at the centre of almost all NASA science activities.

Since then, university researchers have brought innovation and nimbleness to hardware development, have exercised tender loving care of space instruments and have provided a necessary antidote to government stagnation.

Universities have been a fertile training ground for thousands of space engineers and researchers, who have learned to be creative while sticking to budgets and schedules. This has been shown statistically in an analysis of historical data by David Bearden and his colleagues at the Aerospace Corporation in El Segundo, California, due to be published in September.

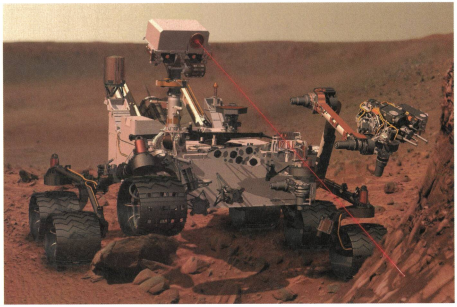
The central and indispensable role of universities in space work is now under immense stress as budgets tighten and NASA withdraws to its centres and core industry contractors. Most people who knew and understood the essential nature of the three-way space partnership are gone. Those who have replaced them in policy and leadership roles may not have realized or absorbed the lessons of the early days.

University labs are being driven out of business. In the recent 'Earth Venture' mission selection, four and a half of the five concepts selected were for missions led by NASA centres. (The half comes from a collaboration between a NASA centre and a university.) Many space hardware-development groups that were thriving as recently as five years ago are now defunct. Those that remain are struggling. Students are finding fewer opportunities for experimentation and are not being trained to do things cheaply in a 'hands-on' fashion. Space research has fallen into a vicious negative-feedback loop.

"University labs are being driven out of business."

For example, all 15 of the high-priority missions recommended for initiation by NASA in the 2007 US National Academies' decadal survey, *Earth Science and Applications from Space*, are being implemented by collaborations between centres and industry. Costs for the first set of missions have ballooned by factors of two to three. According to a mid-term assessment of the survey, in-sourcing work to centres and the use of 'directed' missions rather than competitive PI-class missions were among the reasons for rising costs.

With insufficient funds, missions are being cancelled or delayed. As ageing spacecraft begin to fail, the United States is in grave danger of losing its ability to view Earth from space. Soon, it will be unable to provide decision-makers with the information they need to respond to natural hazards and to the ever-increasing pace of changes that are occurring in the atmosphere, oceans, land surface, cryosphere and ecosystems.



NASA/JPL-CALTECH

NASA's costly Curiosity rover will hopefully touch down on Mars to begin work this week.

THE RIGHT BALANCE

Some things cannot be done in modest PI mode. A dedicated flagship mission is needed to develop, for instance, the multiple-instrument spacecraft necessary for going to the challenging environments of the outer planets. The same is true for large-aperture facility-class astrophysics programmes (including the James Webb Space Telescope).

I am not arguing that NASA should gut its centres. Before my present university position, I was a laboratory director at NASA's Goddard Space Flight Center in Greenbelt, Maryland, from 1987 to 1994. I saw first-hand what immense strengths could be mustered with a critical mass of engineering and science talent. In the current frenzy to cut federal budgets, there is a real danger of losing vital and unique capabilities at centres that have taken years to build and hone.

There have already been staff reductions, such as at the Jet Propulsion Laboratory in Pasadena, California, and the planned budget cuts mean that more losses may be in the offing. We must not allow the navigation, propulsion and communication skills that enable space to be explored to slip through our fingers. If we do, we may never again be able to traverse the rings of Saturn, nor land on an enticing asteroid — nor one day plumb the depths of Europa's oceans.

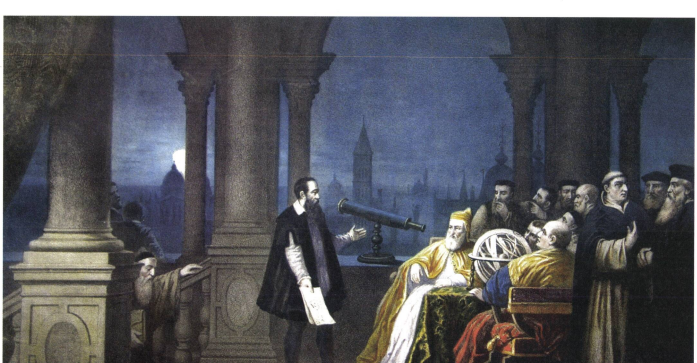
But we must not allow centres to be maintained at exorbitant staffing levels irrespective of cost. Too many institutions employ workers who are performing routine and often-unnecessary functions. I have recently seen dozens of extra managers and engineers assigned to NASA programmes just to give them accounts to charge to. Many centre-led missions are costly because they focus more on maintaining jobs than on getting the biggest scientific bang for the buck.

There is ample support in the space-research community for a more balanced space programme. Many US National Research Council (NRC) reports and decadal surveys have made clear calls for more PI-led missions. And NASA's Explorer programme has used PIs to study focused space physics and astrophysics. Similarly, the Discovery and New Frontiers programmes in its Solar System exploration division are PI missions to study planetary-science issues of moderate scale. But the general trend is away from such missions.

Planning groups at NASA should work with the NRC, the other National Academies and the Office of Management and Budget to shift towards the kind of balanced programme that I advocate here. Allocating a few hundred million more dollars from NASA's $5-billion space-research budget to the PI end of the spectrum could work wonders, in my view.

With government budgets tightening, space research should be revived in universities because they are the best places to foster innovative thinking and to get science done in an affordable way. They are also where we must train the scientists and engineers who will bring an aggressive, nimble mindset to a brighter, future NASA. ■

Daniel N. Baker *is professor of astrophysical and planetary sciences and director of the Laboratory for Atmospheric and Space Physics at the University of Colorado, Boulder, USA. He was chief of the Laboratory for Extraterrestrial Physics at NASA's Goddard Space Flight Center from 1987 to 1994. e-mail: daniel.baker@lasp.colorado.edu*



UNIVERSAL HISTORY ARCHIVE/GETTY

Galileo Galilei, pictured centre, after a painting by H. J. Detouche, is imagined in Giulio Tononi's *Phi* as an explorer of consciousness.

NEUROSCIENCE

A quest for consciousness

Christof Koch marvels at a journey that explains mind–body theory through a fantastical lens.

In the end, consciousness is all that matters. So writes Giulio Tononi, whose stunningly original scientific fantasy, *Phi*, is a distant echo of that great deduction by René Descartes. Tononi, a neuroscientist, psychiatrist and expert on sleep and consciousness, is also that rarest of modern scholars — an idealist. In this category-defying book, he presents his quantitative theory of how brain produces mind as a voyage of discovery imagined for Galileo Galilei.

In Tononi's literary telling of this story, Francis Crick teaches Galileo basic neuroscience. Galileo learns that the brain is the seat of the mind, and that consciousness flees when neurons turn on and off together during deep sleep or seizures, as the pair meet scholars, scientists, doctors and artists from the Enlightenment to the modern era. The vast cast includes Descartes, Nicolaus Copernicus, Charles Darwin, Sigmund Freud, Marcel Proust and, eventually, Alan Turing.

Galileo negotiates some tricky concepts on a road long trodden by neuroscientists and neurologists seeking to track consciousness down to its lair in the brain. Even if we could point to this biophysical mechanism, and those nerve cells, as mediators of the phenomenal experience of red, we would still need to ask — why these particular mechanisms and neurons? Why not others? Historically, the great challenge has been to explain how consciousness emerges from highly organized matter without invoking magic, soul-stuff or exotic physics.

Phi: A Voyage from the Brain to the Soul
GIULIO TONONI
Pantheon: 2012. 384 pp. £19.99, $30

With the advent of Claude Shannon's information theory in the twentieth century, scholars averred a link between information and conscious experience without working out what that might be or could imply. Tononi's integrated-information theory does both. Proceeding from two axioms that are rooted in everyday phenomenal experience, the theory defines a measure (the eponymous Φ) that is associated with every system that consists of causally interacting parts. This measure is high if a system constitutes a single entity above and beyond its parts (integration) and if it is endowed with a large repertoire of discriminable states (information). The more integrated information any system has, the more conscious it is. This framework, couched in a probabilistic language, also captures the unique intrinsic quality of experience — why blue, for example, is more similar to red than to pain or smell.

In *Phi*, this is conveyed through a ▶

series of dazzling thought experiments aided by cameos from Shannon and philosophers Spinoza, Leibniz and Thomas Nagel (the only living person to figure in the book). Through them, Galileo understands how the algebra of integrated information is turned into the geometry of conscious experiences, and how this links to the physiology and the anatomy of the brain.

In the book's final third, Tononi lays out the implications of his theory. He discusses a number of points about consciousness: that it ceases in death and dementia, does not require language or knowledge of self, exists in animals in graded forms and can be present, to some degree, in the fetus.

Hell, Tononi emphasizes, is all in the mind. One of the most chilling characters in *Phi* is the Master, an amalgam of the captain in Franz Kafka's 1914 short story *In the Penal Colony* and the Grand Inquisitor from Fyodor Dostoyevsky's novel *The Brothers Karamazov* (1880). The Master's obsession is creating perfect never-ending pain by manipulating the brain's informational content. In the final chapter, the Mannequin, a stand-in for Mephistopheles, throws up some logical paradoxes before leaving the dying Galileo reunited with his beloved daughter.

"This is a story for grown men, not a consoling tale for children."

Phi is extraordinary. In its appeal to the imagination, it bears some resemblance to Edwin Abbott's *Flatland* novella (1884) or Douglas Hofstadter's *Gödel, Escher, Bach* (Basic Books, 1979). Yet its language is more poetic, and full of cultural references and images — film stills and often modified coloured photos of artworks. Endnotes to each chapter link the allegories and metaphors in the text to the science.

I believe that in the fullness of time, the quantitative framework outlined in *Phi* will prove to be correct. Consciousness is tightly linked to complexity and to information, with profound consequences for understanding our place in the evolving Universe. As Crick says to Galileo, this is a "story for grown men, not a consoling tale for children". ■

Christof Koch *is chief scientific officer at the Allen Institute for Brain Science in Seattle, Washington, and professor of biology and engineering at the California Institute of Technology in Pasadena, California.*
e-mail: christof@alleninstitute.org

Gavin Jantjes' untitled painting depicts a Khoisan girl creating the Milky Way.

ASTRONOMY

Under African skies

Ivan Semeniuk follows the gaze of artists from cultures that have interpreted the heavens for millennia.

Gazing up at a sky full of stars is one of the most universal of human experiences, cutting across cultures and, one imagines, stretching back to the dawn of humanity. Yet artistic depictions of the heavens in popular culture are predominantly European — from Johann Bayer's engravings of the constellations in his 1603 star atlas *Uranometria* to the swirling brilliance of Van Gogh's 1889 painting *The Starry Night*.

An exhibition at the US National Museum of African Art, part of the Smithsonian Institution in Washington DC, may help to change that. It showcases a range of contemporary and historical pieces by African artists. All are connected in one way or another to the Sun, Moon or stars.

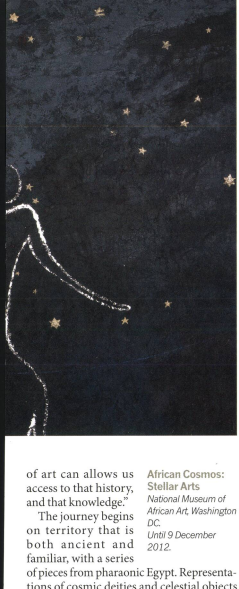
African Cosmos: Stellar Arts was sponsored in large part by the government of South Africa. The country was selected this year, along with Australia, to host the Square Kilometre Array, which will be the world's largest radio telescope; that association adds to the sense of interplay between the scientific and the spiritual that weaves its way through the exhibition. The show seamlessly bridges the centuries, uniting pieces as diverse as traditional moon masks from Côte d'Ivoire and *Trembling Field*, an interactive sculpture by South African Karel Nel. Nel is resident artist with the Cosmic Evolution Survey, a project that focuses on a two-square-degree field of the sky to see how the Universe has changed over time.

"Africa has a long and rich history of keen observation of the heavens," says the exhibition's curator, Christine Mullen Kreamer. "Works

Figures from Central Africa bear lunar patterns.

PHOTOGRAPH: F. KHOURY

METROPOLITAN MUSEUM OF ART



of art can allows us access to that history, and that knowledge."

The journey begins on territory that is both ancient and familiar, with a series of pieces from pharaonic Egypt. Representations of cosmic deities and celestial objects such as the bright star Sirius are reminders of the night sky's prominent role in the rituals and beliefs of civilizations along the Nile. The exhibition goes on to leap across the Sahara and forward in time. Far more exotic to non-African eyes are items that date back only a century or so: a bowl and lid from Nigeria representing the domains of Earth and sky, or a Dogon stool from Mali, which depicts human ancestral figures descending from the heavens to populate the land below.

One of the more striking of the contemporary works, an untitled painting by South African artist Gavin Jantjes, playfully reverses the theme of genesis from above. Based on a Khoisan myth from southern Africa, it depicts the story of a girl dancing around a fire. She throws glowing embers high into the night, thereby creating the Milky Way, the dominant feature of the southern sky.

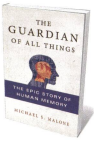
So might the creative sparks tossed skywards from such an exhibition serve to illuminate a continent's worth of artistic achievement and potential. ■

African Cosmos: Stellar Arts
National Museum of African Art, Washington DC.
Until 9 December 2012.

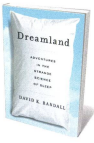
Ivan Semeniuk *is* Nature*'s chief of correspondents in Washington.*

Books in brief

Curious Behavior: Yawning, Laughing, Hiccupping, and Beyond
Robert R. Provine HARVARD UNIVERSITY PRESS *246 pp. £18.95 (2012)*
How can farting, sneezing and other marginal biological realities illuminate humanness? Neuroscientist Robert Provine turns an evolutionary lens on everything from the gross to the faintly improper. The 'contagiousness' of yawning, for instance, hints at the roots of empathy and herd behaviour. Burping and farting were involved in the development of speech, says Provine. And tickling may play a part in our early understanding that we are distinct beings (you can't tickle yourself). An exercise in 'small science' — some of it speculative, all of it fascinating.

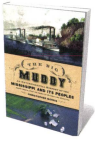


The Guardian of All Things: The Epic Story of Human Memory
Michael S. Malone ST MARTIN'S PRESS *304 pp. £18.99 (2012)*
Memory is a kind of relay, with each generation passing its torch on to the next — creating a conduit for thought and civilization through the eons. In his evocative book, technology writer Michael Malone traces that history from the brain's evolution and the development of speech and writing to advances in recording, the rise of technology and the shifts in ownership of memory from the tribal elect to the masses. The book is packed with gems, including a passage on the twelfth century, when Greek and Arabic science infused Europe, filling its libraries and helping to seed its universities.



Dreamland: Adventures in the Strange Science of Sleep
David K. Randall NORTON *304 pp. £17.99 (2012)*
Sleep occupies us for one-third of our lives. So insomnia, nightmares, deprivation and other aspects of bad sleeping are an obsession for thousands. The tipping point for journalist David Randall was sleepwalking into a wall. Astonished by a specialist's admission of ignorance about the condition, Randall set out to uncover research and shine a light into some dark corners. The entertaining result covers plenty of territory, from the medieval habit of dividing nightly sleeps to the link between vacuum cleaners and sleep apnea. Along the way, Randall picks up the basics for crafting a healthy snooze.

Air: The Restless Shaper of the World
William Bryant Logan NORTON *416 pp. £13.99 (2012)*
Arboriculturist William Bryant Logan follows *Oak* (Norton, 2005) and *Dirt* (Norton, 2007) with this splendid exploration of the "floating world" of air — our planet's invisible skin. Starting with the tornadoes that hit New York in 2010, he both warns of and celebrates the often turbulent and dangerous action of atmosphere. Logan delivers vast amounts of science with brevity and elegance, and is as breezy describing the billion tonnes of dust that blow from African deserts to fertilize the Amazon as he is discussing the echolocation skills of some people with sight impairment.



The Big Muddy: An Environmental History of the Mississippi and Its Peoples, from Hernando de Soto to Hurricane Katrina
Christopher Morris OXFORD UNIVERSITY PRESS *336 pp. £22.50 (2012)*
Seven years ago this month, Hurricane Katrina triggered massive flooding in the valley of the Mississippi. The environmental backstory of the catastrophe is as rich as river sediment, and historian Christopher Morris takes us through 500 years of it. The valley's metamorphosis from vast wetland staked out by France and Spain to a patchwork of development — drained swamp, levees, deforestation, industry and poor urban planning — is powerfully recounted.

Correspondence

Budget cuts leave US science lagging

As a student of science, I'm thrilled that scientists at CERN, Europe's particle-physics lab, have proved the existence of the Higgs boson and advanced our understanding of the Universe. But as an American, I'm somewhat saddened. Had congressional budget-cutters been less short-sighted two decades ago, the Higgs boson might have been discovered by a US-led team instead of by a European consortium. On 4 July, no less.

In 1993, Congress cancelled funding for the Superconducting Super Collider near Waxahachie, Texas, after sinking US$2 billion into an 87-kilometre particle accelerator that promised to establish the United States as the leader in physics research. Two years later, funding was approved by CERN to build the Large Hadron Collider near Geneva, Switzerland.

US science is facing a growing threat from a well-funded anti-science movement, abetted by those corporations and politicians opposed to any research that conflicts with their own vested interests.

Apathy towards basic research in the United States is coupled with an increasing reluctance to invest in science projects that do not have a foreseeable pay-off. But let's not forget that the pioneers of quantum mechanics in the 1900s — Niels Bohr, Albert Einstein and Erwin Schrödinger — were unable to offer any practical ideas about commercial uses for the subatomic particles, quarks and leptons they were bringing to light at the time. However, if you are reading this on a computer, tablet or smart phone, you have quantum mechanics to thank.

An estimated 30–35% of today's US gross domestic product is based on inventions derived from quantum theory, from semiconductors in computer chips and lasers in compact-disc players to magnetic resonance imaging in hospitals and much more.

If it doesn't want to fall behind, the United States should be following the lead of other nations that are investing in science and technology to benefit their economies.

William J. Richards *Hall Institute of Public Policy, New Jersey, USA.*
wrichards@hallnj.org

Improve access to sanitation in China

We hope that last month's raising of drinking-water standards in China will help to speed up improvements in the country's sanitation. As in India, sanitation remains inadequate for a rapidly developing country (*Nature* **486**, 185; 2012).

In 2010, 477 million people in China (36% of the population) did not have access to improved sanitation such as a ventilated pit latrine or a flush toilet piped to a sewer system (WHO/UNICEF *Progress on Drinking Water and Sanitation*, 2012). There are national disparities as well, with 74% of people having improved access to sanitation in urban areas in 2010, but only 56% in rural areas. Provision of sanitation facilities for disabled people is sparse.

China's growing population and urbanization make sewage treatment a particular challenge. Although about 73% of urban sewage is treated (*China Statistical Yearbook on Environment*; 2010), more than 95% of waste water in rural areas drains untreated into rivers and lakes (X. Sun *et al. Chinese Agr. Sci. Bull.* **26**, 384–388; 2010).

The country has now increased its surveillance of freshwater pollution so that the new drinking-water standards can be met. This should catalyse the government into investing more in nationwide sanitation improvements.

Hong Yang, Jim A. Wright *University of Southampton, UK.*
hongyanghy@gmail.com
Stephen W. Gundry *University of Bristol, UK.*

Better lives, not just contraceptives

Last month's London Summit on Family Planning, hosted by the Bill & Melinda Gates Foundation and the UK government's Department for International Development, has been hailed as a resounding success. A total of US$2.6 billion was pledged to provide 120 million women and girls in developing countries with access to family-planning services by 2020. In measuring the success of this welcome campaign, the delivery of social change should also be taken into account.

The hosts emphasize that results will be rapid and quantifiable, for example in terms of the number of contraceptives supplied. But reducing unwanted pregnancies requires other improvements in women's lives, such as better education for girls and reduced child mortality (J. Drèze and M. Murthi *Popul. Dev. Rev.* **27**, 33–63; 2001), outreach by community-health workers and women's empowerment (see, for example, go.nature.com/bpjgma), and quality family-planning programmes. Such factors are harder to quantify.

Focusing simply on what can be measured encourages short-term, narrow interventions rather than broader, longer-term strategies. For instance, value-for-money criteria make it tempting to sidestep national health-care systems, when supporting these is crucial to the delivery of appropriate technologies in developing countries.

Devi Sridhar *University of Oxford, UK.*
devi.sridhar@wolfson.ox.ac.uk
Karen Grépin *New York University, USA.*

Bat deaths from wind turbine blades

You suggest that wind turbines kill bats as a result of air-pressure changes when they fly through the wake of a spinning blade (the barotrauma hypothesis). However, this is likely to be only a minor cause of bat deaths (*Nature* **486**, 310–311; 2012).

The barotrauma hypothesis has been criticized as based on erroneous interpretations of bat injuries (K. E. Rollins *et al. Vet. Pathol.* **49**, 362–371; 2012). Evidence from bat carcasses shows that blunt-force trauma from the spinning blades is a much more common killing mechanism (see also S. M. Grodsky *et al. J. Mammal.* **92**, 917–925; 2011).

We hope that this finding will be useful in mitigating the effects of wind turbines on bat mortality.

Angelo Capparella, Sabine Loew *Illinois State University, Normal, Illinois, USA.*
apcappar@ilstu.edu
David K. Meyerholz *University of Iowa, Iowa City, USA.*

Giants all around — apart from the squid

A smile would have crossed the late Andrew Huxley's face at your description of his "experiments on the axon of the giant squid" (*Nature* **486**, 10–11; 2012). Huxley was giant, the axon was giant, but the squid were quite average in size.

Jonathan C. Horton *University of California, San Francisco, California, USA.*
hortonj@vision.ucsf.edu

CONTRIBUTIONS

Correspondence may be sent to **correspondence@nature.com** after consulting the guidelines at **go.nature.com/cmchno**. Alternatively, readers may comment online: **nature.com/nature**.

RESEARCH

NEW ONLINE

Papers published this week at Nature.com/nature

NEWS & VIEWS A tiny bacterial genome brings the cellular computer a little closer **p.40**

LETTER A 365-million-year-old insect fills a yawning gap in the fossil record **p.82**

LETTER Antisense therapy rescue in a mouse model of muscular dystrophy **p.111**

PUBLISHED 29 JULY 2012

DRUG DISCOVERY

JMJ demethylase inhibitors

Jumonji (JMJ) family histone demethylases are key regulators of transcription, although the importance of their enzymatic activity — as opposed to their structural role — is unclear. JMJD3 demethylases are specific for histone H3 trimethylated at Lys27 (H3K27me3) and are involved in the inflammatory response and other physiological functions. Here a structure-guided small molecule and chemoproteomics approach is used to discover a catalytic site inhibitor selective for the H3K27me3-specific JMJ subfamily. The inhibitor reduces lipopolysaccharide-induced production of pro-inflammatory cytokines in human primary macrophages. This study shows the importance of JMJ demethylase activity and suggests that small-molecule inhibitors of JMJ enzymes could have therapeutic applications.

A selective jumonji H3K27 demethylase inhibitor modulates the proinflammatory macrophage response
L Kruidenier *et al.*
doi:10.1038/nature11262

NEUROSCIENCE

Protocadherins mediate dendritic self-avoidance in the mammalian nervous system
J L Lefebvre
doi:10.1038/nature11305

Structural plasticity and dynamic selectivity of acid-sensing ion channel–spider toxin complexes
I Baconguis & E Gouaux
doi:10.1038/nature11375

STEM CELLS

Neuronal circuitry mechanism regulating adult quiescent neural stem-cell fate decision
J Song *et al.*
doi:10.1038/nature11306

STRUCTURAL BIOLOGY

Structure of a RING E3 ligase and ubiquitin-loaded E2 primed for catalysis
A Plechanovová *et al.*
doi:10.1038/nature11376

PUBLISHED 1 AUGUST 2012

VIROLOGY

Live HIV-infected cells pictured

This study provides a glimpse of the behaviour of HIV-infected cells in a living animal, revealing unanticipated events. Using multiphoton intravital microscopy, Thomas Murooka *et al.* visualize HIV-1 infected T cells within the lymph nodes of humanized mice. They observe reduced cell motility and the formation of long membrane tethers on infected cells. Treating infected mice with a lymphocyte egress inhibitor prevents the spread of HIV-1 in to the circulation, although inhibition persisted only during treatment.

HIV-infected T cells are migratory vehicles for viral dissemination
T T Murooka *et al.*
doi:10.1038/nature11398

MATERIALS SCIENCE

Fast multishell alloy transistors

In the continuing drive to improve and miniaturize transistors, the microelectronics industry has adopted three-dimensional electronic gate structures. Another way to improve transistors is to use semiconductor materials that have higher electron mobility than silicon, although this presents significant fabrication challenges. Katsuhiro Tomioka *et al.* combine these approaches; they grow with high precision vertical, six-sided core-multishell indium-gallium-arsenide nanowires with an all-surrounding gate on a silicon substrate. The authors show superior transistor performance, with excellent on/off switching behaviour and fast operation.

A III–V nanowire channel on silicon for high-performance vertical transistors
K Tomioka *et al.*
doi:10.1038/nature11293

CANCER

Cancer stem cells in solid tumours

Conclusive evidence for the existence of cancer stem cells (CSCs) — cells that can maintain tumour growth — in solid tumours has previously been obtained only in transplantation experiments. Two papers published online this week present evidence of the importance of CSCs. Using genetic lineage tracing, Gregory Driessens *et al.* clonally trace tumour cells *in vivo* in a mouse model of papilloma. The cells in these benign lesions mirror the clonal hierarchy organization of healthy tissue. A small fraction of highly proliferative papilloma cells sustains a slower-cycling transient cell population that gives rise to the tumour cells that form the bulk of the papilloma. On progression to squamous-cell carcinoma, there is a switch to mainly self-renewing divisions and limited differentiation, which is suggestive of CSCs. In a mouse model of glioblastoma, Jian Chen *et al.* use a GFP reporter protein that is expressed only in neural stem cells to trace cancer cells in an endogenous setting. The protein labels a small subset of glioma cells that are less proliferative than the bulk of the tumour cells. However, the GFP^+ cells were responsible for regrowth of the tumours after treatment with the cytotoxic drug temozolomide. Selective ablation of the GFP^+ cells combined with temozolomide was more effective in arresting tumour growth. GFP^+ cells may therefore be CSCs, and targeting both this population and the more proliferative bulk of cancer cells could help to improve therapeutic outcome.

Defining the mode of tumour growth by clonal analysis
G Driessens *et al.*
doi:10.1038/nature11344

A restricted cell population propagates glioblastoma growth after chemotherapy
J Chen *et al.*
doi:10.1038/nature11287

Resolving the stem-cell debate
R J Gilbertson; T A Graham
doi:10.1038/nature11480

NEUROSCIENCE

Lrp4 is a retrograde signal for presynaptic differentiation at neuromuscular synapses
N Yumoto *et al.*
doi:10.1038/nature11348

BRIEF COMMUNICATIONS ARISING

Arising from 'Grid cells without theta oscillations in the entorhinal cortex of bats' by M M Yartsev et al. Nature ***479***, *103–107 (2011)*

Models of grid cells and theta oscillations
C Barry *et al.*
doi:10.1038/nature11276

***Reply* —** M M Yartsev *et al.*
doi:10.1038/nature11277

PALAEONTOLOGY

An insect to fill the gap

A complete insect fossil from the Devonian period has long been sought. The finding of a candidate may improve our patchy understanding of when winged insects evolved. SEE LETTER P.82

WILLIAM A. SHEAR

Insects are, in terms of species number, the most successful group of animals ever to have lived. But their evolutionary origins are a source of controversy, and will continue to be so until the fossil record finally yields up unequivocal evidence of insect beginnings. On page 82 of this issue, Garrouste *et al.*[1] claim to have found precisely this. Although it can hardly be described as well preserved, the fossil shows a six-legged thorax, long single-branched antennae, triangular jaws and a 10-segmented abdomen (see Fig. 2 of the paper[1]). Insects are the only known arthropods (joint-legged invertebrate animals) with this anatomical combination, allowing the authors to make a strong case for the fossil's insectan identity.

The 8-millimetre-long fossil, which the authors named *Strudiella devonica*, was found in a small rock slab excavated at a quarry in Belgium. *Strudiella* is dated to approximately 370 million years old, which places it late in the Devonian period (Fig. 1). This was the time when terrestrial ecosystems were first assembling from their aquatic progenitors[2] — the first forests were established and the earliest four-legged vertebrates were crawling out from freshwater pools onto land. So far, only suggestive traces of insects have been found in rocks of this age. The famous Rhynie chert, a sedimentary deposit in Scotland that is about 402 million years old, contains fossils of collembolans[3], a class of animal that contains today's ubiquitous springtails, and which is regarded as closely related to insects. The Rhynie chert has also yielded a pair of jaw fossils called *Rhyniognatha*, which may be from an advanced, winged insect[4]. In New York state, some fossilized scraps of characteristic cuticle and the framework of a single compound eye, perhaps from a primitive, wingless insect, have been found in 385-million-year old rocks[5]. But these fragments more or less complete the picture of all that is known of insects at this crucial time in Earth's history.

There have also been some false alarms. For example, the fossilized head of a wingless insect found[6,7] in Canadian strata somewhat older than the New York deposits is almost certainly a contaminant — a much more recent or contemporary insect lodged in a crack in the rocks. And *Leverhulmia mariae*, also from Scottish chert near Rhynie, could have been an insect, a close relative, or neither — it seems to have too many legs to be easily classified[8].

So although its age makes it too late to be an insect ancestor, or even the earliest insect, *Strudiella* is nonetheless of great potential significance as the oldest complete insect fossil yet found. This is the first and primary point speaking to its importance.

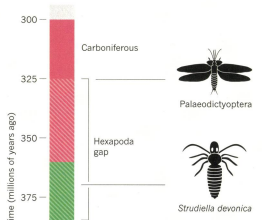


Figure 1 | Winged beginnings. The fossil record provides ample examples of winged insects from around 325 million years ago, such as the order Palaeodictyoptera from the Carboniferous period. However, there is little evidence of insect evolution before this time; only a handful of fossils, including 402-million-year-old *Rhyniella praecursor*, which appears similar to extant collembolan arthropods, have been found. But none of these few examples comes from the period between 385 million and 325 million years ago, which is referred to as the Hexapoda gap (striped region) because of the lack of insect evidence. Now, however, Garrouste *et al.*[1] report the finding of *Strudiella devonica*, a fossil dated to 370 million years ago that shows multiple anatomical features that are characteristic of insects.

We can perceive only what the fossil record permits us to perceive. From our current viewpoint, the diversification of insects and of our own terrestrial vertebrate ancestors seems to have occurred in two evolutionary bursts[9]. Between 425 million and 385 million years ago, both groups probably originated and underwent an initial evolutionary radiation as they began occupying the newly available subaerial realm. There then follows a long period, called Romer's gap (360 million to 345 million years ago) for the vertebrates and the longer Hexapoda gap for insects (385 million to 325 million years ago), during which few, if any, fossils of these groups can be found (Fig. 1). Then, with apparent suddenness, an explosive appearance of many new forms takes place in the second round of diversification. For the insects, large winged species of the major groups (mayflies, proto-dragonflies and others, including extinct types) show up, seemingly without precursors. The insects were off and running on their way to world domination.

These gaps, and the two bouts of evolution that they create, may or may not be real. There is evidence that a period of low atmospheric oxygen concentration coincided with the gap period, and this could have suppressed the rate of appearance of novel anatomy[10]. But a more parsimonious explanation is simply that we have not yet found the right rock formations to reveal fossils that would fill in the gaps. For example, most of the exposed strata for this period in Europe and North America are of marine, not land, origin.

This brings us to the second reason for the importance of *Strudiella* — it is dated to a time smack in the middle of the Hexapoda gap (Fig. 1). According to Garrouste *et al.*, this significantly narrows the gap. And if, as the authors suggest, the fossil came from the young stage of an animal that would have had wings as an adult, their finding would mean that winged insects originated much earlier than fossils have heretofore told us, and that

the sudden appearance of many winged kinds around 325 million years ago is deceptive. It would also suggest that the *Rhyniognatha* fossils could indeed be the mandibles of a winged insect, and that the diversification of winged species could have taken place at a much more leisurely pace, over some 45 million years.

Considering the crucial role of insects in present-day ecology, the number of people engaged in studying their fossil history is dismally small. Furthermore, current specialists focus mostly on events from the Mesozoic period — the 'Age of Dinosaurs' — which began some 70 million years after the time of the first known winged insect fossils, or on even more recent amber-preserved insects, which are largely indistinguishable from living forms[11]. The beginnings of the insects are to be found in rocks much older even than those that enclosed *Strudiella*, but almost no one is looking for them. The paltry few insect fossils contemporary with *Strudiella* — and indeed *Strudiella* itself — were serendipitous, not deliberate, finds, and the Hexapoda gap still looms large. ■

William A. Shear *is in the Department of Biology at Hampden-Sydney College, Hampden-Sydney, Virginia 23943, USA. e-mail: wshear@hsc.edu*

1. Garrouste, R. *et al. Nature* **488,** 82–85 (2012).
2. Shear, W. A. & Selden, P. A. In *Plants Invade the Land. Evolutionary & Environmental Perspectives* (eds Gensel, P. G. & Edwards, D.) 29–51 (Columbia Univ. Press, 2001).
3. Greenslade, P. J. & Whalley, P. E. S. In *Proc. 2nd Int. Semin. Apterygota* (ed. Dallai, R.) 319–323 (Univ. Siena, 1986).
4. Engel, M. S. & Grimaldi, D. A. *Nature* **427,** 627–630 (2004).
5. Shear, W. A. *et al. Science* **224,** 492–494 (1984).
6. Labandeira, C. C., Beall, B. S. & Hueber, F. M. *Science* **242,** 913–916 (1988).
7. Jeram, A. J., Selden, P. A. & Edwards, D. *Science* **250,** 658–661 (1990).
8. Fayers, S. R. & Trewin, N. H. *Palaeontology* **48,** 1117–1130 (2005).
9. Labandeira, C. C. & Sepkoski, J. J. *Science* **261,** 310–315 (1993).
10. Ward, P., Labandeira, C. C., Laurin, M. & Berner, R. A. *Proc. Natl Acad. Sci. USA* **103,** 16818–16822 (2006).
11. Grimaldi, D. A. & Engel, M. S. *Evolution of the Insects* (Cambridge Univ. Press, 2005).

EARTH SCIENCE

The balance of the carbon budget

Careful analysis reveals that the global uptake of anthropogenic carbon dioxide emissions by carbon sinks has doubled during the past 50 years — but the fractions of this absorbed by land and by sea remain unclear. SEE LETTER P.70

INGEBORG LEVIN

Precisely quantifying the fate of man-made carbon dioxide is vital for reliably estimating future atmospheric CO_2 levels and the contribution of this greenhouse gas to global climatic change. There has been much debate about whether currently active carbon sinks are likely to change drastically in the near future[1], or might already have weakened during the past few decades[2,3]. On page 70 of this issue, Ballantyne and colleagues[4] clarify matters. By calculating the worldwide increase in the atmospheric CO_2 burden during the past 50 years from precise global observations, and by carefully accounting for anthropogenic carbon sources and their uncertainties, they find that total global carbon sinks have not declined during this period. Rather, they have increased more than twofold — although fossil-fuel CO_2 emissions have risen almost fourfold[5,6] over this time.

Roughly half of the anthropogenic CO_2 emissions caused by the burning of fossil fuels and by land-use change (such as deforestation) are currently taken up by the oceans or re-enter the terrestrial biosphere — for example, through afforestation or increased biomass production by plants. These carbon sinks have been efficiently damping increases in anthropogenic CO_2 (Fig. 1), but will not necessarily continue to do so. Possible mechanisms that could result in reduced carbon uptake include changes in carbon chemistry and mixing in the oceans[7], and potential feedback from land ecosystems such as increased decomposition of soil organic matter caused by rising global temperatures[8].

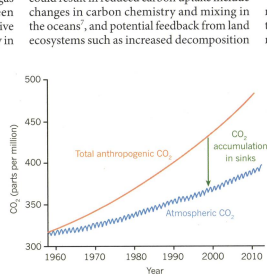


Figure 1 | Partial absorption of anthropogenic carbon dioxide. The graph compares atmospheric levels of carbon dioxide since the 1960s (measured at Mauna Loa, Hawaii[11]) with the levels that would have occurred as a result of the accumulation of anthropogenic CO_2 emissions in the absence of carbon sinks (as calculated by Ballantyne *et al.*[4]). The amount of CO_2 that has accumulated in sinks is represented by the difference between the two curves (green arrow).

To assess variations in carbon sinks over the past 50 years, Ballantyne *et al.* used a strikingly simple approach: they calculated the annual changes in the global atmospheric CO_2 inventory from a worldwide network of long-term observations. By subtracting the annual total amount of anthropogenic CO_2 emissions from these changes, they quantified the net CO_2 uptake by the land and oceans each year. This fundamental strategy reminds us that from time to time we should step back and carefully consider the basic inputs and outputs of the carbon cycle at a global level. Although similar work has been done before[2,9], Ballantyne *et al.* go one step further by quantifying an increase in global CO_2 uptake as worldwide emissions are increasing.

To determine the uncertainties of their estimated carbon budget, the authors combined the most comprehensive, high-accuracy measurements of atmospheric CO_2 available with three inventories of estimated global fossil-fuel CO_2 emissions, and with three inventories of emissions estimated to have been caused by land-use change. The inventories had been made independently of each other, and were obtained using different methods. Bringing together all these data allowed the authors to reliably calculate the errors of their different budget terms. Their finding that the net carbon sink is increasing is therefore robust, and proves that ocean and land sinks have not both decreased in the last half century. How the total sink was divided between oceans and land in the past, and whether the oceans' CO_2-uptake rate has decreased, as suggested by some modelling studies[2], remain open questions.

Why is it so important to know where half of the carbon that we are

currently emitting into the atmosphere actually goes? One reason involves the variable sustainability of carbon sinks. It makes a big difference whether the extra carbon emitted is stored in reservoirs such as the deep oceans, where it could stay for hundreds or thousands of years, or whether it is taken up by the growth of new forests, where it would stay for only a few years or decades before being returned to the atmosphere (either as respired CO_2 or when wood from the forests is used to produce energy). Another equally important reason is the need to understand the processes responsible for carbon uptake, because this knowledge will allow reliable predictions to be made of future atmospheric CO_2 abundance.

This brings us back to the big question posed by Ballantyne and colleagues' work: if CO_2 uptake has increased, where is all the carbon going? Have we overlooked any major carbon-sink processes, and, if so, do we have the right observational strategies in place to detect and quantify these sinks?

We certainly do not have enough precise, consistent, long-term measurements of carbon fluxes and stocks on land, in the oceans or in the atmosphere at all relevant scales — the data that form the backbone of our current understanding of the carbon cycle. For example, the monitoring of changes in carbon uptake by the Southern Ocean requires more precise and highly comparable measurements to be made in that region. The same is true for land areas, such as permafrost regions, that could undergo large changes as a result of changing climatic conditions. Satellite observations may help to provide data for regions for which we have no *in situ* measurements, but at present they are not able to deliver the calibrated, long-term and consistent observations required to quantify global carbon-budget changes. We must therefore set up and maintain a comprehensive framework to monitor all carbon compartments *in situ*. By combining data about bottom-up ocean and land fluxes with those from an atmospheric observational network, and integrating the information at regional, continental and global scales, we should be able to obtain urgently needed answers about carbon sinks.

Besides better global coverage of observations, more-accurate numbers on CO_2 emissions from energy production and land-use change are required. Although such data seem reliable for developed countries, those from emerging economies — the regions that are taking over the lead in energy consumption and CO_2 emissions — make estimates of CO_2 inventories increasingly uncertain[5,10]. Improvements in, and full transparency of, carbon accounting would help not only in negotiations aimed at reducing CO_2 emissions, but also in precisely budgeting natural CO_2 fluxes. This is true both on the global scale studied by Ballantyne *et al.* and for regional fluxes where carbon sinks are active. ■

Ingeborg Levin *is at the Institut für Umweltphysik, Heidelberg University, Heidelberg D-69120, Germany.*
e-mail: ingeborg.levin@iup.uni-heidelberg.de

1. Friedlingstein, P. *et al. J. Clim.* **19,** 3337–3353 (2006).
2. Sarmiento, J. L. *et al. Biogeosciences* **7,** 2351–2367 (2010).
3. Zhao, M. & Running, S. W. *Science* **329,** 940–943 (2010).
4. Ballantyne, A. P., Alden, C. B., Miller, J. B., Tans, P. P. & White, J. W. C. *Nature* **488,** 70–72 (2012).
5. Marland, G., Hamal, K. & Jonas, M. *J. Ind. Ecol.* **13,** 4–7 (2009).
6. http://cdiac.ornl.gov/trends/emis/prelim_2009_2010_estimates.html
7. Gruber, N. *Phil. Trans. R. Soc. A* **369,** 1980–1996 (2011).
8. Piao, S. *et al. Nature* **451,** 49–52 (2008).
9. Knorr, W. *Geophys. Res. Lett.* **36,** L21710 (2009).
10. Guan, D., Liu, Z., Geng, Y., Lindner, S. & Hubacek, K. *Nature Clim. Change* advance online publication http://dx.doi.org/10.1038/nclimate1560 (2012).
11. www.esrl.noaa.gov/gmd/ccgg/trends

DRUG DISCOVERY

Kill the messenger where it lives

A mutant repeating DNA sequence produces a toxic RNA molecule that causes the neuromuscular disorder myotonic dystrophy type 1. An 'antisense' therapy that targets this RNA in cell nuclei shows promise in mice. SEE LETTER P.111

PETER K. TODD & HENRY L. PAULSON

Myotonic dystrophy type 1 is a devastating inherited disorder for which effective treatments are lacking. It is characterized by myotonia (delayed relaxation of muscles after contraction), weakness, cardiac arrhythmias, diabetes and cognitive changes. Patients carry a mutant *DMPK* gene that contains a greatly expanded DNA repeat sequence within a non-coding region[1–3]. When the mutant gene is expressed, it yields a messenger RNA molecule that seems to be toxic to cells such as muscle fibres[4] — which means that the inheritance of a single mutant copy of *DMPK* from either parent is sufficient to cause disease. On page 111 of this issue, Wheeler and colleagues[5] describe the successful use of antisense oligonucleotides (synthetic RNA-like fragments that bind to a target RNA) to correct the molecular and physiological features of the disease in a mouse model.

The mutation in *DMPK* elicits the toxicity associated with myotonic dystrophy type 1 (DM1) through at least two mechanisms[4]. Usually, *DMPK*-encoded mRNA is synthesized in the cell's nucleus and efficiently exported to the cytoplasm, where it is translated into protein (Fig. 1a). However, owing to its expanded repeat, the mutant RNA forms a hairpin-shaped structure that binds to members of the MBNL family of proteins, which regulate RNA splicing, a process by which immature mRNAs are cut up and reassembled before translation (Fig. 1b). As a result, the mutant RNA and the proteins form aggregates ('foci') in the nucleus, and MBNL activity is decreased. In a second mode of action, the mutation somehow triggers increased levels of a different splicing protein, CELF1. The opposing effects of the *DMPK* mutation on CELF1 and MBNL activities disrupt RNA splicing and so lead to disease[4].

A potential strategy for treating DM1 is to 'silence' the mutant *DMPK* using antisense oligonucleotides (Fig. 1c). Because patients carry a functional copy of the gene in addition to the altered one, oligonucleotides could be designed that would bind selectively to the mutant RNA and mark it for degradation — while leaving translation of the functional RNA unaffected. This strategy can be tested in a mouse model of DM1, in which the expanded repeat from a mutant *DMPK* has been added to an unrelated gene, the expression of which is easy to track in muscle[6]. Indeed, it has been shown that direct injection of a specific type of antisense oligonucleotide (morpholino oligonucleotides) into the muscles of such mice reduced the toxicity of the mutant RNA, although systemic delivery proved inefficient[7].

Wheeler *et al.* used 'gapmer' antisense oligonucleotides, which contain chemical modifications at their ends that make them more stable. Moreover, gapmer oligonucleotides include a central sequence that, when bound to its target RNA, promotes target cleavage by the enzyme RNase H. The authors administered gapmer oligonucleotides to DM1 model mice by subcutaneous injection, and observed a robust and sustained decrease in the concentration of the toxic RNA in muscle, even when applying relatively low doses. This decrease was associated with improvement in a wide range of disease features — for example, the authors observed loss of nuclear foci, release of MBNL proteins from nuclear aggregates,

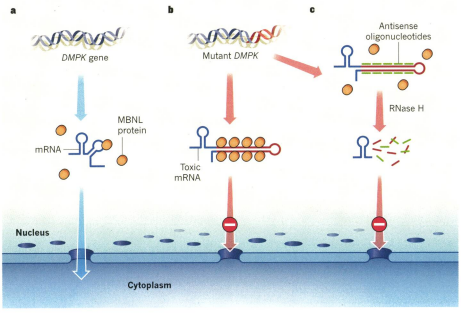


Figure 1 | How to silence a toxic RNA. **a**, Most messenger RNAs, such as that encoded by the *DMPK* gene, are processed by splicing proteins and then rapidly exported from the nucleus into the cytoplasm, where they are translated into proteins. MBNL splicing proteins interact with many mRNAs, including the *DMPK* mRNA. **b**, Expansion of a repeat region in *DMPK* mRNA causes myotonic dystrophy type 1 by sequestering MBNL proteins and retaining them in the nucleus, thereby affecting the splicing and expression of many cellular RNAs. **c**, Wheeler and colleagues[2] describe the successful use of antisense oligonucleotides (short RNA-like molecules) to ameliorate the disease's symptoms in a mouse model. The oligonucleotides bind to the mutant RNA and selectively induce its destruction in the nucleus by the enzyme RNase H.

and correction of RNA splicing defects and the resultant myotonia. Astonishingly, the most effective oligonucleotides continued to confer some benefit up to a year after treatment had been discontinued.

Animal tests of RNA-directed therapies for muscle diseases such as DM1 have had limited success so far[7–9]. So, why this apparent breakthrough? It comes down to the therapy's probable site of action: the nucleus. Most mRNAs are synthesized and spliced in the nucleus, then rapidly exported to the cytoplasm. But gapmer oligonucleotides induce degradation of RNA by RNase H, which is enriched in the nucleus and almost absent from the cytoplasm. Wheeler and colleagues, and a second research group working independently[9], reasoned that the nuclear retention of expanded-repeat RNAs could make them good targets for RNase H. Consistent with this idea, the authors describe how oligonucleotides designed to target RNAs that are rapidly exported to the cytoplasm were ineffective at decreasing their expression in muscle. By contrast, oligonucleotides targeting a nuclear RNA (the long non-coding RNA *Malat1*) demonstrated similar efficacy to that seen when targeting the expanded-repeat RNA.

What are the implications of Wheeler and colleagues' results? They inspire optimism that previous challenges faced by researchers looking at antisense oligonucleotide therapies for DM1 and other neuromuscular diseases are surmountable — although significant hurdles remain regarding safety and delivery to affected tissues other than skeletal muscle, such as the heart and brain. The authors' findings also suggest that gapmer-based strategies might be suitable for the treatment of other disorders caused by expansions of repeated DNA sequences (such as amyotrophic lateral sclerosis and frontotemporal dementia[10,11]), provided that the mutant RNA tends to remain in the cell's nucleus longer than the normal RNA. Furthermore, appropriately designed gapmer oligonucleotides may aid researchers in defining the functions of specific nuclear non-coding RNAs, some of which have key roles in regulating gene expression.

However, as promising as these findings are for the prospect of DM1 therapeutics, they also serve as a cautionary tale for the applications of antisense oligonucleotides. First, given our limited understanding of the roles of nuclear non-coding RNAs and the likelihood that their sensitivity to this technology is enhanced, care must be taken in oligonucleotide design to avoid potentially deleterious off-target effects. Second, developing similarly potent therapies for target mRNAs that are rapidly exported from the nucleus may require the use of oligonucleotides that do not act through RNase H. Third, therapeutic success in a mouse model is still a long way from effective application in humans. However, the path to success now seems clearly visible. ■

50 Years Ago

The importance of lunar natural resources for the future of space exploration can scarcely be exaggerated. Lunar resources will not only play an important part in the establishment of a lunar base by providing life support materials and vehicle fuels but will also be an important, and perhaps a limiting, factor in the logistics of interplanetary space exploration. Certainly, only in the most cursory exploration of the solar system could be conducted using either existing or planned propulsion systems so long as the rocket vehicles must lift all their fuel from the surface of the Earth. A lunar fuel source, on the other hand, would provide an extremely convenient low-gravity refuelling station in space.

From *Nature* 4 August 1962

100 Years Ago

A note bearing on the much-debated question of the age of the earth is given in the Proceedings of the Tokyo Mathematico-physical Society by S. Suzuki. The calculation refers to the time taken for the present crust of the earth to solidify. A result is obtained on the supposition that the heat of fusion liberated by the solidification of the crust supplies the heat lost by radiation, and it is further assumed that the effect of the curvature of the earth's surface may be neglected. According to these hypotheses the calculated time varies between 30 and 300 million years, according to the kind of rock (gneiss, basalt, or granite) assumed in the calculations. The difficulty is, of course, our imperfect knowledge of the experimental data on which the conclusions are based.

[Editor's note: Latest estimates give Earth's age as 4.5 billion years.]

From *Nature* 1 August 1912

Peter K. Todd *and* **Henry L. Paulson** *are in the Department of Neurology, University of Michigan, Ann Arbor, Michigan 48109, USA.*
e-mails: henryp@umich.edu; petertod@umich.edu

1. Brook, J. D. *et al. Cell* **68,** 799–808 (1992).
2. Harley, H. G. *et al. Nature* **355,** 545–546 (1992).
3. Buxton, J. *et al. Nature* **355,** 547–548 (1992).
4. Renoux, A. J. & Todd, P. K. *Prog. Neurobiol.* **97,** 173–189 (2012).
5. Wheeler, T. M. *et al. Nature* **488,** 111–115 (2012).
6. Mankodi, A. *et al. Science* **289,** 1769–1773 (2000).
7. Wheeler, T. M. *et al. Science* **325,** 336–339 (2009).
8. Mulders, S. A. *et al. Proc. Natl Acad. Sci. USA* **106,** 13915–13920 (2009).
9. Lee, J. E., Bennett, C. F. & Cooper, T. A. *Proc. Natl Acad. Sci. USA* **109,** 4221–4226 (2012).
10. Renton, A. E. *et al. Neuron* **72,** 257–268 (2011).
11. DeJesus-Hernandez, M. *et al. Neuron* **72,** 245–256 (2011).

ALZHEIMER'S DISEASE

A protective mutation

A rare gene variant has been found that decreases the peptide deposition seen in the brains of people with Alzheimer's disease. The mutation may also slow the normal cognitive decline that occurs with age. SEE LETTER P.96

BART DE STROOPER & THIERRY VOET

Although our understanding of the biological processes causing Alzheimer's disease remains hazy, one clear defining feature of the disease is an accumulation in the brain of deposits of amyloid-β peptides. These form following cleavage of the amyloid precursor protein (APP), and mutations in the *APP* gene can affect this cleavage and/or the biophysical properties of the peptides[1], leading to their aggregation into toxic peptide assemblies[2,3] that propagate in the brain and promote neurodegeneration[4,5]. All *APP* mutations identified so far increase levels of these toxic amyloid-β species, but might *APP* mutations exist that act in the opposite way, by reducing peptide accumulation and protecting against neurodegeneration? On page 96 of this issue, Jonsson *et al.*[6] report finding the proverbial needle in a haystack: a rare variant of the *APP* gene that protects against Alzheimer's disease.

The mutation causes an amino-acid substitution in APP (alanine to threonine; A673T) close to the site at which the protein is cleaved by the enzyme β-secretase (Fig. 1). The authors show that fewer amyloid-β (Aβ) peptides are generated in cultured cells that express this gene variant. It is worth noting that a different amino-acid substitution at exactly the same site (alanine to valine; A673V) can cause Alzheimer's disease[7], emphasizing the pathological relevance of this amino-acid residue.

Fascinatingly, Jonsson and colleagues also found that, in a cohort of people over the age of 80, those who were heterozygous for the A673T variant (one of their two copies of the *APP* gene was mutated) performed better in a test of mental capacity than did control subjects. The authors derive from this observation the rather bold conclusion that the identified variant not only protects against Alzheimer's disease but also against the normal mild cognitive decline that is associated with old age[6].

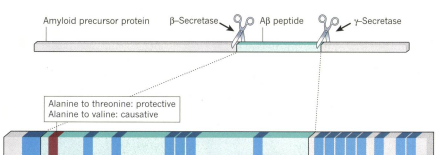


Figure 1 | Amyloid precursor protein and Alzheimer's disease. The amyloid precursor protein (APP) is cleaved by the enzymes β-secretase and γ-secretase into amyloid-β (Aβ) peptides (scissor symbols represent the cleavage sites). An accumulation of Aβ peptides is seen in the brains of patients with Alzheimer's disease, and multiple mutations that cause changes in the amino-acid sequence of APP (blue bars) are associated with the disease[4,15]. By contrast, Jonsson and colleagues[6] have identified a mutation in the *APP* gene that reduces Aβ accumulation and that protects against Alzheimer's disease. The mutation results in an alanine-to-threonine amino-acid substitution close to the β-secretase cleavage site (red bar). Interestingly, an alanine-to-valine substitution at the same site can cause Alzheimer's disease[7].

In their search for rare *APP* mutations with significant effect on the risk of developing Alzheimer's disease, Jonsson and colleagues used a unique collection of genotypic, medical and genealogical records of Icelandic people gathered by the company deCODE Genetics. The authors screened the full genome sequences of 1,795 Icelanders for variants in and outside the *APP* gene and used computational methods to predict these variant sequences in approximately 370,000 compatriots. This combination of sequence and relationship information[8,9] enabled the authors to test the association of the *APP* variants they identified with common late-onset Alzheimer's disease. They found that the protective A673T variant was significantly more common in a group of over-85-year-olds without Alzheimer's disease (the incidence was 0.62%) — and even more so in cognitively intact over-85-year-olds (0.79%) — than in patients with Alzheimer's disease (0.13%). Although these findings were largely based on predicted genotypes, the authors demonstrated the accuracy of their calculations by subsequently testing for the A673T gene variant in thousands of the individuals they studied.

Jonsson and colleagues also used *in vitro* studies to show that the mutated APP protein undergoes about 40% less cleavage by β-secretase compared with wild-type APP. Together, β-secretase and γ-secretase are responsible for the cleavage of Aβ peptides from APP (Fig. 1); the enzymes are currently being assessed as targets for the treatment of Alzheimer's disease[10], and the finding of reduced Aβ accumulation following impaired β-secretase cleavage is encouraging for this line of inquiry. At first glance, one is tempted to draw the exciting conclusion that a lifelong, moderate lowering of Aβ is all that is needed to postpone Alzheimer's disease and to maintain good cognition. Indeed, researchers have long been struggling with the tantalizing question of to what extent levels of Aβ must be lowered to be effective in the treatment or prevention of Alzheimer's disease[11].

However, it remains to be seen how these *in vitro* observations translate in the human brain, where the protective A673T variant is expressed together with a wild-type copy of the gene. If the translation is linear, then the 40% decrease in Aβ levels observed *in vitro* would result in a 20% decrease in Aβ in people carrying the mutation. It would be extremely instructive to test this prediction using blood, cerebrospinal fluid or cells from individuals carrying the mutation.

Despite this promising prospect, the question remains as to whether a reduction in Aβ explains the protective effect of the gene variant identified by Jonsson and colleagues. Here, it is worth keeping in mind a previously identified Alzheimer's-disease-causing mutation, the A673V *APP* gene variant[7]. This mutation increases Aβ generation but causes dementia only in people in which both gene copies are mutated, not just one. It also affects not only the amount, but also, and importantly for this discussion, the biophysical properties of the Aβ that is generated. It seems that the mutated protein interacts with wild-type Aβ to prevent the generation of toxic Aβ assemblies. Given that another mutation at the same site in the APP protein also affects the aggregation properties of Aβ peptides[12], the possibility that Jonsson and colleagues' A673T mutation exerts its protective effects by altering Aβ aggregation should be considered. This more qualitative concept of Aβ toxicity contrasts with the idea that only an increase in Aβ levels can cause disease — and evidence supporting this insight is rapidly mounting[13].

Further work is certainly needed to verify whether the A673T mutation protects against age-related cognitive decline. Jonsson *et al.* report that A673T carriers perform better in cognitive tests than do control subjects, but one wonders whether this can be confirmed by other measurements of cognitive function, and whether confounding factors complicate the interpretation of the reported result. For example, although there were no known cases of Alzheimer's disease in the control population, many other conditions, such as Parkinson's disease or depression (which were not excluded in this assessment), can also negatively affect mental capacity. So Jonsson and colleagues' proposal[6] that "Alzheimer's disease may represent the extreme of the age-related decline in cognitive function" may yet prove to be a premature interpretation of their findings.

Nevertheless, the identification of a protective *APP* gene variant is certainly exciting, and it will be interesting to watch for the identification of other protective gene variants, for example, mutations in the gene encoding β-secretase that might inhibit its expression or its proteolytic activity. As with many genetic findings, more years of hard work will be needed to assess the clinical and therapeutic implications of such findings. But if the preliminary — and quite spectacular — conclusions of Jonsson *et al.* regarding the mechanism of action of the A673T mutation, and its implications for cognition, can be confirmed, then a lifelong suppression of Aβ production by as little as 20% may one day become the 'fountain of youth' for the brain. ■

Bart De Strooper *and* **Thierry Voet** *are in the Center for Human Genetics and Leuven Research Institute for Neurodegenerative Disorders, KU Leuven, 3000 Leuven, Belgium.* **B.D.S.** *is also at the VIB Center for the Biology of Disease, Leuven.*
e-mail: bart.destrooper@cme.vib-kuleuven.be

1. De Strooper, B. *Physiol. Rev.* **90**, 465–494 (2010).
2. Haass, C. & Selkoe, D. J. *Nature Rev. Mol. Cell Biol.* **8**, 101–112 (2007).
3. Benilova, I., Karran, E. & De Strooper, B. *Nature Neurosci.* **15**, 349–357 (2012).
4. Eisenberg, D. & Jucker, M. *Cell* **148**, 1188–1203 (2012).
5. Stöhr, J. *et al. Proc. Natl Acad. Sci. USA* **109**, 11025–11030 (2012).
6. Jonsson, T. *et al. Nature* **488**, 96–99 (2012).
7. Di Fede, G. *et al. Science* **323**, 1473–1477 (2009).
8. Kong, A. *et al. Nature Genet.* **40**, 1068–1075 (2008).
9. Marchini, J., Howie, B., Myers, S., McVean, G. & Donnelly, P. *Nature Genet.* **39**, 906–913 (2007).
10. De Strooper, B., Vassar, R. & Golde, T. *Nature Rev. Neurol.* **6**, 99–107 (2010).
11. Karran, E., Mercken, M. & De Strooper, B. *Nature Rev. Drug Discov.* **10**, 698–712 (2011).
12. Meinhardt, J. *et al. Protein Sci.* **16**, 1214–1222 (2007).
13. Chávez-Gutiérrez, L. *et al. EMBO J.* **31**, 2261–2274 (2012).
14. Wilquet, V. & De Strooper, B. *Curr. Opin. Neurobiol.* **14**, 582–588 (2004).
15. Goate, A. & Hardy, J. *J. Neurochem.* **120**, 3–8 (2012).

B.D.S. declares competing financial interests.
See go.nature.com/uwtjbg for details.

QUANTUM OPTICS

Strongly interacting photons

A fine marriage between atomic and optical physics has produced a medium that is transparent to single photons but opaque to multiple photons. The finding heralds the development of devices such as single-photon switches. SEE LETTER P.57

THAD G. WALKER

Can photons be made to interact strongly with each other? Until recently, materials with nonlinear optical properties could mediate photon–photon interactions that were weak at best. These weak interactions have previously been artificially enhanced using devices known as optical cavities[1], which make the photons repeat their encounters thousands to millions of times. On page 57 of this issue, Peyronel *et al.*[2] demonstrate a new material in which single photons propagate freely, but interact so strongly with each other that when just two photons are present one is quickly absorbed. The result opens up the possibility of realizing concepts such as single-photon switches, deterministic photon-based quantum logic, and quantum gases of strongly interacting photons*.

We have known since the dawn of quantum physics a century ago that light consists of particles, called photons, of energy hf, where h is Planck's constant and f is the light's frequency. Photons usually interact extremely weakly with each other, but strongly with the charged particles that comprise matter. In most materials, the optical response is linear — a beam comprised of many photons scatters and moves from place to place in the same way that single photons do. Inside nonlinear materials, however, the optical response is altered when multiple photons are present. The motion of a particular photon depends on the properties — most notably the number — of other photons in its vicinity. Until recently, however, available nonlinear materials required large numbers of photons to be present in order for them to noticeably affect each other. Peyronel *et al.*[2] combined several recent developments in atomic and optical physics to produce a novel nonlinear medium that is transparent to single photons yet opaque to multiple photons (Fig. 1).

The largest nonlinear optical effects achievable in atoms occur when a light field renders the atoms transparent. Consider a sample of atoms that have three energy levels: a ground state E_g, an excited state E_e and an intermediate level E_r (see Fig. 1b of the paper[2]). Photons of frequency f_1 that are directed into the sample and obey the Bohr equation, $hf_1 = E_e - E_g$, will normally be absorbed. However, when a strong 'control' laser of frequency f_2 is also shone on the sample, the atoms become transparent to frequency f_1 if the condition $h(f_1 + f_2) = E_r - E_g$ is satisfied. This electromagnetically induced transparency (EIT; ref. 3) puts the atoms and photons into collective excitations called polaritons. In Peyronel and colleagues' experiment, the control laser changes the transmission of the atomic gas from essentially zero to 60%. The EIT condition is extremely sensitive: photons that obey it are transmitted with high probability, whereas those that violate it are absorbed normally. This is basically an optically controlled switch[4].

When the upper level r is a state of large

*This article and the paper under discussion[2] were published online on 25 July 2012.

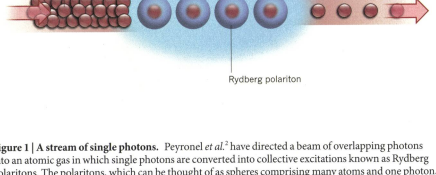


Figure 1 | A stream of single photons. Peyronel *et al.*[2] have directed a beam of overlapping photons into an atomic gas in which single photons are converted into collective excitations known as Rydberg polaritons. The polaritons, which can be thought of as spheres comprising many atoms and one photon, strongly absorb additional photons. On exiting the gas, the polaritons are converted back to individual, non-overlapping photons.

principal quantum number n (a Rydberg state[5]), the EIT condition can be easily violated by weak interactions between the atoms. For an n of about 100, a single Rydberg atom will cause a violation of the EIT condition for all other atoms within a 'blockade radius' of 10 micrometres. This Rydberg blockade produces record nonlinearities, as shown recently by Adams and colleagues[6], and has been used to entangle neutral atoms separated by micrometre-scale distances[7,8]. A Rydberg polariton can be thought of as a 10-μm sphere containing many ground-state atoms and one Rydberg atom — or, equivalently, many atoms and one photon. Should other photons enter a volume already occupied by a Rydberg polariton, the blockade effect causes a violation of the EIT condition, so the photons are absorbed rather than transmitted. Note that if the atom density is low, as in previous experiments[6], the absorption probability may still be small.

The final, essential ingredient needed to generate strong photon–photon interactions at the two-photon level is an atomic cloud of such high density that when two or more photons enter a blockade volume, all but one are absorbed within that volume, leaving a single Rydberg polariton. This 'photon blockade' is the novelty of Peyronel and colleagues' study. Their experiment reveals that a multi-photon incident light beam is converted, within a few micrometres, into a beam of single photons, with a small (less than 0.09) probability that two photons will leave the atomic gas at the same time. Interestingly, even though their sample is large enough for several Rydberg polaritons to coexist, the authors find that (and explain why) only one photon at a time is found within the entire sample.

An exciting feature of this experiment is that there are several clear avenues towards improving the properties of the medium. Cooler, denser atomic gases and lasers that have a narrower frequency range would improve the EIT transmission to nearly 100% and reduce the overlap of photons from the single-photon source. A looming challenge is to reconfigure the experiment so that the two-photon nonlinearity delays rather than absorbs excess photons[6]. This type of nonlinearity, which preserves the number of photons, would be extremely useful for quantum-information purposes.

In one respect, Peyronel and colleagues have demonstrated a quality single-photon light source that has a rate of emission in the megahertz regime, as Dudin and Kuzmich have shown[9] using a related approach. The key capability of this experiment[2] — engineering strong photon–photon interactions at the two-photon level — should also lead to various other new possibilities. For example, single-photon switches, photon detectors of high quantum efficiency, and non-destructive photon detection can easily be foreseen as extensions of this work. The physics of strongly interacting photons has a bright future. ■

Thad G. Walker *is in the Department of Physics, University of Wisconsin-Madison, Madison, Wisconsin 53706, USA.*
e-mail: tgwalker@wisc.edu

1. Birnbaum, K. M. *et al. Nature* **436,** 87–90 (2005).
2. Peyronel, T. *et al. Nature* **488,** 57–60 (2012).
3. Fleischhauer, M., Imamoglu, A. & Marangos, J. P. *Rev. Mod. Phys.* **77,** 633–673 (2005).
4. Harris, S. E. & Yamamoto, Y. *Phys. Rev. Lett.* **81,** 3611–3614 (1998).
5. Saffman, M., Walker, T. G. & Molmer, K. *Rev. Mod. Phys.* **82,** 2313–2363 (2010).
6. Pritchard, J. D., Weatherill, K. J. & Adams, C. S. Preprint at http://arxiv.org/abs/1205.4890 (2012).
7. Wilk, T. *et al. Phys. Rev. Lett.* **104,** 010502 (2010).
8. Isenhower, L. *et al. Phys. Rev. Lett.* **104,** 010503 (2010).
9. Dudin, Y. O. & Kuzmich, A. *Science* **336,** 887–889 (2012).

SYSTEMS BIOLOGY

A cell in a computer

The small genomes of some bacteria could provide the first complete understanding of a biological system. A new computer model brings this goal closer, by calculating every process in a dividing *Mycoplasma* cell.

MARK ISALAN

It has long been a dream in biology to push reductionism to the limit: to describe a cell as a set of interacting components and to capture whole-cell behaviour in a computer model. A good model doesn't simply recapitulate the observed behaviours that are fed into it. Rather, the aim is to predict the unknown effect of any novel perturbation or mutation. Such goals are very ambitious because of the challenge of attempting to obtain quantitative information on every one of the cell's gene products and metabolites. Nevertheless, Karr *et al.*[1], writing in *Cell*, present the most comprehensive model of a bacterial cell cycle so far, built on the basis of individual molecules and their relationships. Impressively, the model can predict gene-expression levels and cell-replication times in the challenging context of mutations involving gene deletions.

Mycoplasma genitalium is a urogenital bacterial parasite that has only 525 genes, making it one of the smallest genomes of any independently dividing cell — for comparison, the gut bacterium *Escherichia coli* has around 4,000 genes. Because of their status as one of the 'simplest' cells, *Mycoplasma* species are rapidly becoming the most measured biological systems in history, and full descriptions of their molecular content, in terms of DNA, RNA, protein and metabolites, are available[2–4]. The cells are therefore considered to be the ideal target for whole-cell modelling[5].

What is striking about Karr and colleagues' model is the sheer ambition of its scale and its attention to detail. The authors retrieved (and in some cases retested) more than 1,900 experimentally derived cellular parameters, such as enzymatic reaction rates and protein-binding

affinities, from around 900 publications. They then combined these to make 28 sub-models of cellular processes, such as metabolism, protein translation and DNA replication. They used sub-models so that they could apply the appropriate modelling method for each process. In computational biology, this requirement has been neatly summarized[2] as "Don't model bulldozers with quarks". So the authors combined different modelling techniques involving varying levels of detail, to allow different factors — including dependence on deterministic reactions, known constraints, probability and random variability — to be applied where appropriate.

Crucially, the authors then used a computational trick to join up the sub-models (Fig. 1a). Models calculate variables — numbers that represent varying system states. And variables change according to sets of rules — the equations and parameters used to describe the system. The authors allowed each sub-model to calculate independently the values of a set of 16 variables at a time-step of approximately one second. They then combined these results, which generated a new set of variables, and the process was repeated in a loop. Thus, all the sub-models 'communicated' with one another and the cell's status was constantly updated and recalculated. Although this is an approximation, because in reality all processes happen simultaneously, the end results converged plausibly towards the decision to divide, which the authors assessed as the moment that the bacterial cell membrane 'pinched' together to form two new cells (Fig. 1b).

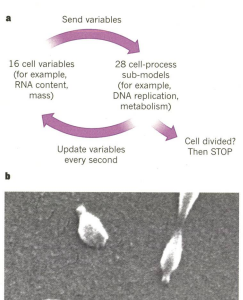

a
Send variables
16 cell variables (for example, RNA content, mass)
28 cell-process sub-models (for example, DNA replication, metabolism)
Update variables every second
Cell divided? Then STOP
b

Figure 1 | Looping calculations to model cell division. **a**, Karr *et al.*[1] have constructed a computer model that attempts to calculate every process in *Mycoplasma genitalium* cells. Their modelling strategy involves 28 independent sub-models of cell processes, each incorporating different methods and levels of detail. The sub-models communicate by combining their calculations for 16 cell variables for approximately one second of the cell's life cycle, and then calculating the next second. The looping process culminates when cell division is induced. **b**, A scanning electron micrograph of *M. genitalium* cells, before (left) and during division (right). Scale bar, 0.5 micrometres.

After some optimization, the model produced estimates of metabolite concentrations, metabolism rates, and messenger RNA and protein levels, that were similar to experimental data. The model also allowed the authors to make several predictions about cell behaviour, including that 90% of the cell's genes will be expressed in the first 2.5 hours of the approximately 9-hour cell cycle. This prediction suggests that the chromosomes are 'explored' rapidly by gene-expression machinery. The key test, however, was whether higher-level system properties, such as the time taken for the cell to replicate itself, would be correctly predicted for bacteria carrying genetic mutations. When the researchers ran the model with each of the 525 genes individually deleted, they found that 284 of the genes are essential for cell survival and 117 are non-essential. These numbers are approximately 80% in agreement with experimental data for gene deletions that have been assessed previously.

The authors also tested the growth rates of 12 of the gene-deleted bacterial strains, and found 8 to be within the limits predicted by the model. In some of these cases, the experiments resolved discrepancies between the model and published growth rates, identifying, for example, a previously undescribed slow-growth mutant.

So, can the authors claim to have recreated a cell in a computer? They themselves say that the model should be compared to the first draft of the human genome, and be considered a work in progress. However, in modelling, discrepancies between predictions and experimental results are the key to improvements — they direct more detailed analyses and model refinement, and ultimately lead to better models. More challenging tests could be imagined. For example, could the model predict synthetic lethal mutants, in which the combination of two gene deletions will kill a cell, although either deletion alone permits survival? Furthermore, any model that attempts to predict phenotypes[7] (biological properties) from genotypes (gene sequences) will be subject to the problem that even genetically identical cells do not always give the same output. For example, random differences in the amount of chaperone proteins can 'buffer' mutations variably[8]. However, the *Mycoplasma* model can track such variability and therefore has the potential to predict these outcomes.

The metaphor of gene networks being connected in wiring diagrams is becoming commonplace and, even though such networks can be non-intuitive[9], they are ideal for computer modelling. Nevertheless, one of the most exciting ideas in studies of gene regulation is that network relationships may not always involve direct molecular interactions. For example, imagine a gene that is required for cell division — if there is low gene expression, the cell will not divide and so gene-expression levels will have longer to accumulate, which can create a feedback loop, or gene expression 'according to need'[10,11]. Fascinatingly, there is already a hint of this in one example from Karr and colleagues' model. They find that cell-cycle time is affected by the concentrations of DNA nucleotides (dNTPs), which are required for DNA replication. When dNTP levels are low, the cycle slows at the point of replication initiation, allowing dNTPs to build up, which then speeds up the rest of the cycle.

M. LLUCH-SENAR

Extrapolating from such findings, I can imagine that similar feedback processes might exist for any cellular factor that contributes indirectly to reducing its own concentration. For example, factors for which transiently low concentrations reduce the activity of pathways for cell division, protein secretion or protein degradation might similarly self-regulate, restoring and buffering themselves over time. So perhaps the most exciting thing about a whole-cell model is that it may allow us to look beyond the direct molecular 'cogs and wheels' that drive biology and into the emergent properties of biological systems. ■

Mark Isalan *is at the EMBL/CRG Systems Biology Research Unit, Centre for Genomic Regulation and UPF, 08003 Barcelona, Spain.*
e-mail: isalan@crg.es

1. Karr, J. R. *et al. Cell* **150,** 389–401 (2012).
2. Güell, M. *et al. Science* **326,** 1268–1271 (2009).
3. Yus, E. *et al. Science* **326,** 1263–1268 (2009).
4. Kühner, S. *et al. Science* **326,** 1235–1240 (2009).
5. Tomita, M. *et al. Bioinformatics* **15,** 72–84 (1999).
6. Goldenfeld, N. & Kadanoff, L. P. *Science* **284,** 87–89 (1999).
7. Jelier, R., Semple, J. I., Garcia-Verdugo, R. & Lehner, B. *Nature Genet.* **43,** 1270–1274 (2011).
8. Burga, A., Casanueva, M. O. & Lehner, B. *Nature* **480,** 250–253 (2011).
9. Isalan, M. *BioEssays* **31,** 1110–1115 (2009).
10. Tsuru, S. *et al. Mol. Syst. Biol.* **7,** 493 (2011).
11. Kashiwagi, A., Urabe, I., Kaneko, K. & Yomo, T. *PLoS ONE* **1,** e49 (2006).

nature

2 August 2012 / Vol 488 / Issue No. 7409

Publication policies, author guidelines and information

This is a summary of *Nature*'s guide to authors, publication policies and other author information, available in full on the web via our main author information page at www.nature.com/nature/authors.

Getting published in *Nature*

Please read www.nature.com/nature/authors/get_published before finalizing any original scientific research manuscripts you intend to submit. This section explains *Nature*'s editorial criteria, and how manuscripts are handled by the editors after submission and before acceptance for publication.

Formatting guide

Please read www.nature.com/nature/authors/gta before finalizing and submitting your contribution. This section provides instructions for preparing, formatting and writing all types of contribution published by *Nature*.

Publication policies

This section, at www.nature.com/authors, describes *Nature*'s conditions of publication, including policies on authorship; competing financial interests; duplicate publication; prepublicity and embargoes; materials and data availability; research on animal (including human) subjects; referees; corrections, retractions, errata and addenda to published papers; the complaints procedure and refutations; and copyright/licence to publish. All contributions, for whatever section of the journal, are considered for publication and published on the condition that authors agree to these policies.

Submissions

These sections, at www.nature.com/nature/authors/submissions, contain our online service for presubmission enquiries and submission and tracking of Articles, Letters and Review-type manuscripts. Authors should read the formatting guide before using this service. Authors who cannot submit online can submit by post, but *Nature* strongly recommends the online service. **Authors must not submit by both methods.** Authors experiencing operational difficulty in using the online submission service can send an enquiry to mtshelp@nature.com, with a brief explanation of the problem.

Editorial contacts and author services

This section, at www.nature.com/nature/about/contact, provides the addresses of *Nature*'s editorial offices; names of editorial staff; and contact details for advice, reprints, permissions, press office (press@nature.com), and other author-related contact information, including postal addresses for mailed submissions. Short biographies of editors are also available at www.nature.com/nature/about/editors.

Final submissions

This section, at www.nature.com/nature/authors/submissions/final, applies only if an Article or Letter has been accepted in principle for publication and the author has been asked by the editor for a final version. Authors will need to upload electronic files of the text, production-quality figures and supplementary information (if any) to our database. Authors must include a completed manuscript checklist and all relevant signed forms and declarations as requested by the editor.

Short version of guide to authors

See www.nature.com/nature/authors for full version. Contributors not used to *Nature*'s format are advised to read some papers in published issues before submission, to familiarize themselves with *Nature*'s style. *Nature*'s main formats for original research are Articles and Letters.

ARTICLES are original reports whose conclusions represent a substantial advance in understanding of an important problem and have immediate, far-reaching implications. They do not normally exceed **5 pages** of *Nature* and have **no more than 50** references. (One page of undiluted text is about 1,300 words.)

Articles have a summary for readers outside the discipline, separate from the main text and containing no references, of up to 150 words. The summary does not contain numbers, abbreviations, acronyms or measurements unless essential. It contains a brief account of the background and rationale of the work, followed by a statement of the main conclusions, introduced by the phrase 'Here we show' or its equivalent.

The Article itself typically has about 3,000 words of text, beginning with up to 500 words of referenced text expanding on the background to the work, followed by a concise, focused account of the findings, and ending with one or two short paragraphs of discussion. The text can contain a few short subheadings consisting of about 20 but no more than 40 characters (including spaces) each.

LETTERS are short reports of original research focused on an outstanding finding whose importance means that it will be of interest to scientists in other fields. They do not normally exceed **4 pages** of *Nature*, and have **no more than 30** references. They begin with a **fully referenced paragraph** of not more than 300 words, aimed at readers in other disciplines. This paragraph contains a summary of the background and rationale for the work, followed by a one-sentence statement of the main conclusions introduced by the phrase 'Here we show' or its equivalent, and ending with two to three sentences describing the implications in context.

The main text which follows is typically up to 1,500 words long. This text can, if the author requires, start with a further brief paragraph of introductory material. A description of the research then follows. Any discussion at the end of the text should be as succinct as possible.

FIGURES in both Articles and Letters are as small as possible: Articles typically have 5 or 6 figures, Letters typically have 3 or 4 figures. Legends and/or methods are additional to the main text. Methods Summaries (which are published in print only) are no longer than 300 words; Methods sections (which are published online) can be up to 1,000 words; legends should not exceed 100 words each and start with a one-sentence title for the figure, followed by a brief explanation of the parts and symbols.

See www.nature.com/nature/authors/gta/others.html for details of other types of contribution to *Nature*, including Brief Communications Arising; Reviews and Perspectives (presubmission enquiries to reviewsandinsights@nature.com); and Correspondence (submit to corres@nature.com).

See Authors and Referees @ NPG (www.nature.com/authors) for details of Nature journals' publication policies, author and referee services, author benefits and for links to Nautilus and Peer to Peer blogs.

ARTICLE

doi:10.1038/nature11213

Novel mutations target distinct subgroups of medulloblastoma

Giles Robinson[1,2,3]*, Matthew Parker[1,4]*, Tanya A. Kranenburg[1,2]*, Charles Lu[1,5], Xiang Chen[1,4], Li Ding[1,5,6], Timothy N. Phoenix[1,2], Erin Hedlund[1,4], Lei Wei[1,4,7], Xiaoyan Zhu[1,2], Nader Chalhoub[1,2], Suzanne J. Baker[1,2], Robert Huether[1,4,8], Richard Kriwacki[1,8], Natasha Curley[1,2], Radhika Thiruvenkatam[1,2], Jianmin Wang[1,9], Gang Wu[1,4], Michael Rusch[1,4], Xin Hong[1,5], Jared Becksfort[1,9], Pankaj Gupta[1,9], Jing Ma[1,7], John Easton[1,4], Bhavin Vadodaria[1,4], Arzu Onar-Thomas[1,10], Tong Lin[1,10], Shaoyi Li[1,10], Stanley Pounds[1,10], Steven Paugh[1,11], David Zhao[1,9], Daisuke Kawauchi[1,12], Martine F. Roussel[1,12], David Finkelstein[1,4], David W. Ellison[1,7], Ching C. Lau[1,13], Eric Bouffet[1,14], Tim Hassall[1,15], Sridharan Gururangan[1,16], Richard Cohn[1,17], Robert S. Fulton[1,5,6], Lucinda L. Fulton[1,5,6], David J. Dooling[1,5,6], Kerri Ochoa[1,5,6], Amar Gajjar[1,3], Elaine R. Mardis[1,5,6,18], Richard K. Wilson[1,5,6,19], James R. Downing[1,7], Jinghui Zhang[1,4] & Richard J. Gilbertson[1,2,3]

Medulloblastoma is a malignant childhood brain tumour comprising four discrete subgroups. Here, to identify mutations that drive medulloblastoma, we sequenced the entire genomes of 37 tumours and matched normal blood. One–hundred and thirty–six genes harbouring somatic mutations in this discovery set were sequenced in an additional 56 medulloblastomas. Recurrent mutations were detected in 41 genes not yet implicated in medulloblastoma; several target distinct components of the epigenetic machinery in different disease subgroups, such as regulators of H3K27 and H3K4 trimethylation in subgroups 3 and 4 (for example, *KDM6A* and *ZMYM3*), and CTNNB1–associated chromatin re–modellers in WNT–subgroup tumours (for example, *SMARCA4* and *CREBBP*). Modelling of mutations in mouse lower rhombic lip progenitors that generate WNT–subgroup tumours identified genes that maintain this cell lineage (*DDX3X*), as well as mutated genes that initiate (*CDH1*) or cooperate (*PIK3CA*) in tumorigenesis. These data provide important new insights into the pathogenesis of medulloblastoma subgroups and highlight targets for therapeutic development.

Medulloblastoma is the most common malignant childhood brain tumour[1]. The disease includes four subgroups (sonic hedgehog (SHH) subgroup, WNT subgroup, subgroup 3 and subgroup 4), defined primarily by gene expression profiling, that show differences in karyotype, histology and prognosis[2]. Studies of genetically engineered mice show that these tumours arise from different cell types: SHH-subgroup medulloblastomas develop from committed cerebellar granule neuron progenitors (GNPs) in *Ptch1*$^{+/-}$ mice[3,4]; WNT-subgroup tumours are generated by lower rhombic lip progenitors (LRLPs) in *Blbp-Cre*;*Ctnnb1*$^{+/lox(Ex3)}$;*Tp53*$^{flx/flx}$ mice[5]; whereas subgroup-3 medulloblastomas probably arise from an undefined class of cerebellar progenitors[6]. The identification of medulloblastoma subgroups has not changed clinical practice. All patients currently receive the same combination of surgery, radiation and chemotherapy. This aggressive treatment fails to cure two thirds of patients with subgroup-3 disease, and probably over-treats children with WNT-subgroup medulloblastoma who invariably survive with long-term cognitive and endocrine side effects[2,7]. Drugs targeting the genetic alterations that drive each medulloblastoma subgroup could prove more effective and less toxic, but the identity of these alterations remains largely unknown.

The genomic landscape of medulloblastoma

To identify genetic alterations that drive medulloblastoma, we performed whole-genome sequencing (WGS) of DNA from 37 tumours and matched normal blood (discovery cohort). Tumours were subgrouped by gene expression (WNT subgroup, $n = 5$; SHH subgroup, $n = 5$; subgroup 3, $n = 6$; subgroup 4, $n = 19$; 'unclassified' (profiles not available), $n = 2$; Fig. 1, Supplementary Figs 1–3 and Supplementary Table 1). Validation of all putative somatic alterations including single nucleotide variations (SNVs), insertion/deletions (indels) and structural variations (SVs) identified by CREST[8], was conducted for 12 tumours using custom capture arrays and Illumina-based DNA sequencing (Supplementary Table 2). Putative coding alterations and SVs were validated in the remaining 25 discovery cohort cases by polymerase chain reaction (PCR) and Sanger-based sequencing. Mutation frequency was determined in a separate 'validation cohort' of 56 medulloblastomas (WNT subgroup, $n = 6$; SHH subgroup, $n = 8$; subgroup 3, $n = 11$; subgroup 4, $n = 19$; unclassified, $n = 12$; Fig. 1 and Supplementary Table 1).

WGS of the discovery cohort detected 22,887 validated or high-quality somatic sequence mutations (SNVs and indels), 536 validated or curated SVs, and 5,802 copy number variations (CNVs; 92%

[1]St Jude Children's Research Hospital, Washington University Pediatric Cancer Genome Project, Memphis, Tennessee 38105, USA. [2]Department of Developmental Neurobiology, St Jude Children's Research Hospital, Memphis, Tennessee 38105, USA. [3]Department of Oncology, St Jude Children's Research Hospital, Memphis, Tennessee 38105, USA. [4]Department of Computational Biology and Bioinformatics, St Jude Children's Research Hospital, Memphis, Tennessee 38105, USA. [5]The Genome Institute, Washington University School of Medicine in St Louis, St Louis, Missouri 63108, USA. [6]Department of Genetics, Washington University School of Medicine in St Louis, St Louis, Missouri 63108, USA. [7]Department of Pathology, St Jude Children's Research Hospital, Memphis, Tennessee 38105, USA. [8]Department of Structural Biology, St Jude Children's Research Hospital, Memphis, Tennessee 38105, USA. [9]Department of Information Sciences, St Jude Children's Research Hospital, Memphis, Tennessee 38105, USA. [10]Department of Biostatistics, St Jude Children's Research Hospital, Memphis, Tennessee 38105, USA. [11]Department of Pharmaceutical Sciences, St Jude Children's Research Hospital, Memphis, Tennessee 38105, USA. [12]Department of Tumour Biology and Genetics, St Jude Children's Research Hospital, Memphis, Tennessee 38105, USA. [13]Texas Children's Cancer and Hematology Centers, 6701 Fannin Street, Ste. 1420, Houston, Texas 77030, USA. [14]The Hospital for Sick Children, 555 University Avenue, Toronto, Ontario M5G 1X8, Canada. [15]The Royal Children's Hospital, 50 Flemington Road, Parkville, Victoria 3052, Australia. [16]Duke University Medical Center, 102382, Durham, North Carolina 27710, USA. [17]The School of Women's and Children's Health, University of New South Wales, Kensington, New South Wales NSW 2052, Australia. [18]Siteman Cancer Center, Washington University School of Medicine in St Louis, St Louis, Missouri 63108, USA. [19]Department of Medicine, Washington University School of Medicine in St Louis, St Louis, Missouri 63108, USA.
*These authors contributed equally to this work.

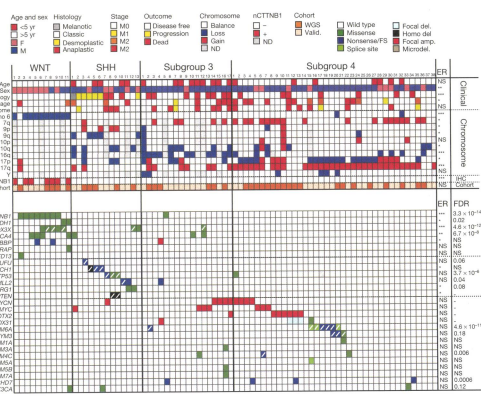


Figure 1 | The genomic landscape of medulloblastoma. Top, clinical, histological, gross chromosomal, nuclear CTNNB1 (nCTNNB1) and cohort (discovery or validation) details of 79 medulloblastomas by subgroup. ER, enrichment. Bottom, genetic alterations detected in 27 genes of particular interest. Colour key at top right. ANOVA (continuous) or Fisher's exact test (categorical) *P* value is shown on right. False discovery rate (FDR) estimates of each mutation are shown on right. Slash indicates loss or mutation of wild-type allele, including X chromosome in males. ***$P < 0.0005$; **$P < 0.005$; *$P < 0.05$; NS, not significant. F, female; M, male. amp., amplification; del., deletion; microdel., microdeletion; valid., validation cohort. ND, not done.

concordant with 6.0 SNP mapping arrays; Supplementary Tables 3–6 and Supplementary Figs 4–7). In all but five tumours with the highest mutation rates, >50% of SNVs were C→T/G→A transitions (Supplementary Fig. 8). The mean missense:silent mutation ratio was 3.6:1 and 40% of all missense mutations were predicted to be deleterious, suggesting a selective pressure for SNVs that affect protein coding (Supplementary Table 5). Global patterns of total SNVs and amplifications varied significantly among medulloblastoma subgroups, even when corrected for age and sex, supporting the notion that these tumours are distinct pathological entities (Fig. 1 and Supplementary Fig. 6). Custom capture-based analysis of the allele frequency of all somatic mutations in 12 medulloblastomas allowed us to predict the ancestry of certain genetic alterations, suggesting that aneuploidy precedes widespread sequence mutation in medulloblastomas with highly mutated genomes (Supplementary Figs 9–11).

Novel CNVs and SVs are rare in medulloblastoma

The repertoire of focally amplified or deleted genes seems to be very limited in medulloblastoma. We detected expected[2] gains of *MYC*, *MYCN* and *OTX2* in subgroups 3 and 4, but no novel recurrent amplifications (Fig. 1, Supplementary Fig. 12 and Supplementary Table 7). In keeping with recent reports[9], high-level amplification of *MYCN* in subgroup-3 sample no. 16 (sample numbering as in Fig. 1) was generated by chromothripsis; although chromothripsis was observed infrequently ($n = 2/37$ of the discovery cohort; Supplementary Fig. 13).

Focal homo- or heterozygous deletions of genes previously implicated in medulloblastoma were also detected (for example, *PTCH1*, *PTEN*; Fig. 1)[10,11] but novel recurrent focal deletions were rare. Three subgroup-4 tumours (nos 11–13) and one unclassified tumour deleted *DDX31*, *AK8* and *TSC1* at chromosome 9q34.14 in concert with *OTX2* amplification, suggesting that these alterations are cooperative ($P < 0.0005$, Fisher's exact test). The breakpoint in this deletion occurs in *DDX31*, and two samples contained a missense mutation (subgroup 4, no. 15) and complex rearrangement (unidentified case SJMB026) in this gene, suggesting that *DDX31* is the target of these alterations (Supplementary Fig. 14).

Over 50% of SVs detected by WGS broke the coding region of at least one gene, but less than 2% ($n = 6/314$, excluding two tumours with excessive SVs) encode potential in-frame fusion proteins (Supplementary Fig. 15); none affect the same gene or signal pathway. Therefore, fusion proteins are likely to be an uncommon transforming mechanism in medulloblastoma.

Although germline mutations in *TP53*, *PTCH1*, *APC* and *CREBBP* predispose to medulloblastoma[11–14], only 23 mutations previously associated with cancer were detected in discovery cohort germ lines. Only one of these—in a known case of Turcot's syndrome—was accompanied by a somatic mutation (germline APC Y935*/somatic deletion; WNT subgroup no. 11; Supplementary Table 8). Thus, inherited forms of medulloblastoma seem to be rare in our cohort.

Novel mutations in medulloblastoma subgroups

Because SVs and CNVs are unlikely to drive most medulloblastomas, we investigated whether recurrent (more than two samples) somatic SNVs and/or indels might target discrete genes and pathways. This analysis identified 49 genes, across all 93 tumours, which were targeted by non-silent, recurrent, somatic mutations; 84% ($n = 41/49$) have not yet been implicated in medulloblastoma (Supplementary Tables 9 and 10). Several of these congregated in disease subgroups and converged on specific cell pathways (Fig. 1, Supplementary Fig. 8 and Supplementary Table 11).

Histone methylation is deregulated in subgroups 3 and 4

The H3K27 trimethyl mark (H3K27me3) represses lineage-specific genes in stem cells[15] (Supplementary Fig. 8). H3K27me3 is written by the polycomb repressive complex 2 (PRC2) that includes the methylase EZH2 (refs 16, 17) and is erased during differentiation by the demethylase KDM6A[18]. As H3K27me3 is erased, chromatin remodellers recruited to H3K4me3 promote differentiation, for example, CHD7 (refs 19, 20). This process is tightly controlled during development and deregulated in cancers; *EZH2* is mutated in lymphomas[21] and upregulated in breast[22] and prostate[23] cancer, while biallelic inactivation of *KDM6A* (chromosome Xp11.2) or *KDM6A* and its paralogue *UTY* (chromosome Yq11), occurs in adult female and male cancers, respectively[24].

Hypergeometric distribution analyses revealed selective mutation of histone modifiers in subgroup-3 and -4 medulloblastomas (Supplementary Table 11). Six subgroup-4, one subgroup-3, and

one unclassified medulloblastoma contained novel inactivating mutations in *KDM6A* (Figs 1 and 2 and Supplementary Figs 8 and 16). The single female with a *KDM6A* splice-site mutation showed a deletion of the second allele that escapes X inactivation[25] (subgroup 4, no. 15), and 57% ($n = 4/7$) of *KDM6A*-mutant male medulloblastomas deleted chromosome Y, compared with only 6% ($n = 3/51$) of male, *KDM6A* wild-type tumours ($P < 0.005$, Fisher's exact test; Fig. 1). Thus, a two-hit model of *KDM6A-UTY* tumour suppression seems to operate in subgroup-4 medulloblastomas. Notably, mutations in six other KDM family members (*KDM1A*, *KDM3A*, *KDM4C*, *KDM5A*, *KDM5B* and *KDM7A*) were detected exclusively in subgroup-3 and -4 tumours, implicating broad disruption of lysine demethylation in these medulloblastomas (Fig. 1, Supplementary Table 11 and Supplementary Fig. 16).

Subgroup-3 and -4 medulloblastomas also gained and overexpressed *EZH2* (chromosome 7q35-34), which writes H3K27me3, and contained novel inactivating mutations in effectors and regulators of the H3K4me3 mark[26] (Fig. 2a and Supplementary Fig. 8). Gain of chromosome 7q was significantly enriched among subgroup-3 and -4 medulloblastomas ($P < 0.005$, Fisher's exact test) and correlated directly with EZH2 expression. Indeed, *EZH2* was the eighth most significantly overexpressed gene on chromosome 7 among subgroup-3 and -4 medulloblastomas that gained chromosome 7q relative to those with diploid chromosome 7 ($P < 0.005$, Bonferroni correction). Nonsense and frameshift mutations were detected in *CHD7* in four subgroup-3 and -4 tumours. *ZMYM3* (chromosome Xq13.1), which participates in a protein complex with KDM1A to regulate gene expression at the H3K4me3 mark[27], was targeted by novel frameshift, nonsense and missense mutations in three male subgroup-4 medulloblastomas. All three tumours with mutations in *ZMYM3* also mutated *KDM6A* (subgroup 4, nos 19, 20) or *KDM1A* (subgroup 4, no. 21), suggesting that these alterations are cooperative. Remarkably, *KDM6A*, *CHD7* and *ZMYM3* mutations were confined to subgroups 3 and 4, and clustered in samples with sub-median EZH2 expression levels (Fig. 2a; $P < 0.05$, Fisher's exact test). These data suggest that subgroup-3 and -4 medulloblastomas retain a stem-like epigenetic state by aberrantly writing (EZH2 upregulation) or preserving (*KDM6A-UTY* inactivation) H3K27me3, or disrupting H3K4me3 associated transcription (*CHD7* and *ZMYM3* inactivation). Indeed, human and mouse subgroup-3 and -4 medulloblastomas contained significantly more H3K27me3 than did WNT- or SHH-subgroup tumours (Fig. 2b). Thus, gain of *EZH2* and loss of *KDM6A* probably maintains H3K27me3 in subgroup-3 and -4 medulloblastomas.

Finally, we looked to see if the differential expression of H3K27me3 among medulloblastoma subgroups reflects ancestral chromatin marking in the progenitors that generate these tumours (Fig. 2b). Relatively low levels of H3K27me3 were detected in LRLPs and committed GNPs, which generate WNT- and SHH-subgroup medulloblastomas, respectively[3–5], potentially explaining why mutations that preserve this epigenetic mark are absent from these tumours. We recently showed that subgroup-3 medulloblastomas arise from a rare fraction of cerebellar progenitors[6]. We are currently investigating whether these progenitors are found among the H3K27me3-positive cells seen in the external germinal layer (Fig. 2b).

Novel mutations in WNT-subgroup medulloblastomas

WNT-subgroup medulloblastomas contained mutations in epigenetic regulators that are different to those seen in subgroup-3 and -4 disease. CTNNB1, the principal effector of the WNT pathway, forms a transcription factor with the T-cell factor/lymphoid enhancer factor (TCF/LEF)[28]. The carboxy terminus of CTNNB1 then recruits a series of protein complexes that remodel chromatin and promote transcription at WNT-responsive genes (Supplementary Fig. 8). These include: histone acetyltransferases (for example, CREBBP and TRRAP–TIP60 complexes)[28,29]; ATPases of the SWI/SNF family (for example, SMARCA4)[30]; and the mediator complex that coordinates RNA polymerase II placement (for example, MED13)[31]. As expected, >70% ($n = 8/11$) of WNT-subgroup medulloblastomas contained mutations that stabilize CTNNB1 (Fig. 1 and Supplementary Fig. 8; $P < 0.0001$, Fisher's exact test)[32,33]. A single subgroup-3 case (no. 5)

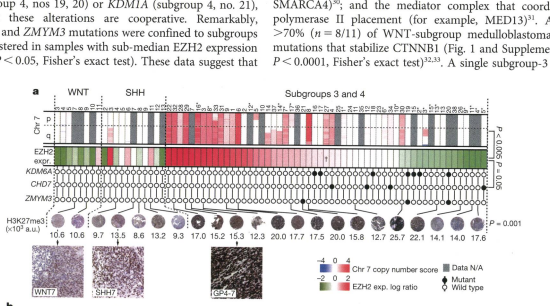


Figure 2 | Deregulation of H3K27me3 in subgroup-3 and -4 human and mouse medulloblastoma. a, Top row, SNP profiles of chromosome 7 (Chr 7) copy number in medulloblastomas (samples as Fig. 1; asterisk indicates subgroup-3 cases). Second row, expression of EZH2. Subgroup-3 and -4 tumours are ordered left to right by expression level, dagger indicates median expression point (Bonferroni-corrected *P* value of EZH2 expression versus chromosome 7 gain). Third row, mutation status of *KDM6A*, *CHD7* and *ZMYM3* (*P* value, Fisher's exact test mutations versus EZH2 expression). Fourth row, H3K27me3 immunohistochemistry (numbers indicate colorimetry, *P* value ANOVA). GP4-7 indicates case subgroup-4, no.7. a.u., arbitrary units. N/A, not available. **b**, H3K27me3 expression in mouse *Blbp-Cre;Ctnnb1*$^{+/lox(ex3)}$;*Tp53*$^{lox/lox}$ (WNT-subgroup), *Ptch1*$^{+/-}$;*Tp53*$^{-/-}$ (SHH-subgroup) and *Myc;Nmyc1*$^{+/-}$ (subgroup-3) medulloblastomas (right) and developing hindbrain (left). High-power views of E14.5 LRL (i) and upper rhombic lip (URL) (ii). EGL, external germinal layer; IGL, internal granule layer. Scale bar, 50 μm. White arrows in P7 cerebellum (CB) pinpoint H3K27me3 cells in the EGL.

also showed a mutation in *CTNNB1*, but this mutation has not been reported in cancer, did not upregulate nuclear CTNNB1 (Fig. 1) and is of unclear relevance. Remarkably, six WNT-subgroup medulloblastomas showed mutations in chromatin modifiers that are recruited to TCF/LEF WNT-responsive genes by CTNNB1 (Fig. 1 and Supplementary Fig. 8). Four WNT-subgroup tumours contained heterozygous missense mutations in the helicase domain of *SMARCA4* ($P < 0.002$, Fisher's exact test), two samples, including one with a *SMARCA4* mutation (no. 5), contained nonsense mutations in *CREBBP* (WNT-subgroup enrichment, $P < 0.02$, Fisher's exact test), and missense mutations in *TRRAP* and *MED13* were detected in a single WNT-subgroup medulloblastoma each. Thus, in addition to stabilization of CTNNB1, the development of WNT-subgroup medulloblastoma may require disruption of chromatin remodelling at WNT-responsive genes.

A small number of WNT-subgroup medulloblastomas lack mutations in *CTNNB1* or *APC*, suggesting that alternative mechanisms drive aberrant WNT signals in these tumours. Three WNT-subgroup medulloblastomas in our series contained wild-type *CTNNB1* (nos 1, 10 and 11; Fig. 1). Sample no. 11 inactivated *APC* as the sole case of Turcot's syndrome in our study, but this tumour and sample no. 10 also contained novel missense mutations in *CDH1* (R63G, V329F; WNT-subgroup enrichment, $P < 0.05$, Fisher's exact test; Fig. 1). CDH1 sequesters CTNNB1 at the cell membrane[34], and mutations that disrupt this interaction promote WNT signalling in adult cancers[35,36]. The functional consequences of CDH1(R63G) and CDH1(V329F) remain to be determined, but their restriction to WNT-subgroup tumours, mutual exclusivity with *CTNNB1* mutations, and adjacency to residues mutated in breast cancer (http://www.sanger.ac.uk/genetics/CGP/cosmic/), suggest they might promote aberrant WNT signals in medulloblastoma.

We showed previously in mice that mutant Ctnnb1 initiates WNT-subgroup medulloblastoma by arresting the migration of LRLPs from the embryonic dorsal brainstem to the pontine grey nucleus (PGN)[5]. Therefore, to test whether disruption of CDH1 might substitute for mutant CTNNB1 in medulloblastoma, we used short hairpin (sh)RNAs to knockdown *Cdh1* in embryonic day (E)14.5 mouse LRLPs (Fig. 3a–c). Deletion of Cdh1 expression upregulated Tcf/Lef-mediated gene transcription in LRLPs and more than doubled their self-renewal capacity (Fig. 3b). Furthermore, *in utero* electroporation of LRLPs with *Cdh1* shRNAs impeded their migration from the dorsal brainstem to the PGN with an efficiency similar to that of mutant Ctnnb1 (Fig. 3d, e; see Supplementary Methods). These data support the hypothesis that *CDH1* suppresses the formation of WNT-subgroup medulloblastoma by regulating WNT-signals in LRLPs.

WNT-subgroup medulloblastomas were also enriched for novel, recurrent somatic missense mutations in the DEAD-box RNA helicase *DDX3X* at chromosome Xp11.3 ($P < 0.0001$, Fisher's exact test; Fig. 1). DDX3X regulates several critical cell processes including chromosome segregation[37], cell cycle progression[38], gene transcription and translation[39]. Previously reported cancer-associated mutations in *DDX3X* disrupt the ATPase activity of the protein, but seven of eight mutations identified in our series clustered in the DEAD-box domain (Supplementary Information and Supplementary Fig. 8). Structural modelling predicts that these mutations interfere with nucleic acid binding, possibly altering specificity and/or affinity for RNA substrates, rather than inactivating DDX3X (Supplementary Figs 17–22). Indeed, the wild-type allele of *DDX3X* that escapes X inactivation[25] was retained by two of three *DDX3X*-mutant female medulloblastomas, and knockdown of *Ddx3x* halved the self-renewal rate of mouse LRLPs, suggesting that this protein is important for the proliferation and/or maintenance of the LRLP lineage (Fig. 3b).

To understand better the role of DDX3X in WNT-subgroup medulloblastoma, we used our *in utero* migration assay to assess the impact of *Ddx3x* shRNAs, mutant $Ddx3x^{T275M}$ (identified in WNT-subgroup sample no. 9), or mutant $Ddx3x^{G325E}$ (WNT sample no. 8) on LRLPs. Remarkably, although *Ddx3x* shRNAs were expressed abundantly in E14.5 brainstem cells within 48 h of electroporation, ≤0.5% of *Ddx3x*-shRNA-positive cells were present by postnatal day (P)1, confirming the critical importance of this gene to maintain the LRLP lineage (Fig. 3d, e). In contrast, mice electroporated with either mutant $Ddx3x^{T275M}$ or $Ddx3x^{G325E}$ consistently contained

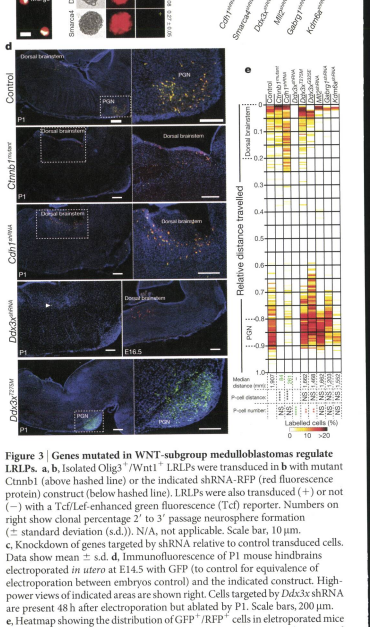


Figure 3 | Genes mutated in WNT-subgroup medulloblastomas regulate LRLPs. a, **b**, Isolated $Olig3^+/Wnt1^+$ LRLPs were transduced in **b** with mutant Ctnnb1 (above hashed line) or the indicated shRNA-RFP (red fluorescence protein) construct (below hashed line). LRLPs were also transduced (+) or not (−) with a Tcf/Lef-enhanced green fluorescence (Tcf) reporter. Numbers on right show clonal percentage 2′ to 3′ passage neurosphere formation (± standard deviation (s.d.)). N/A, not applicable. Scale bar, 10 μm. **c**, Knockdown of genes targeted by shRNA relative to control transduced cells. Data show mean ± s.d. **d**, Immunofluorescence of P1 mouse hindbrains electroporated *in utero* at E14.5 with GFP (to control for equivalence of electroporation between embryos control) and the indicated construct. High-power views of indicated areas are shown right. Cells targeted by *Ddx3x* shRNA are present 48 h after electroporation but ablated by P1. Scale bars, 200 μm. **e**, Heatmap showing the distribution of GFP^+/RFP^+ cells in eletroporated mice at P1. Median distance migrated by cells and *P* values of migration distance and cell number relative to controls is shown. $^{****}P < 0.00005$; $^{***}P < 0.0005$; $^{**}P < 0.005$; $^{*}P < 0.05$. Red and green text reports significant increase or decrease, respectively, relative to control.

~50% more labelled cells at P1 than did controls, although these cells migrated normally (Fig. 3d, e and data not shown). Thus, mutations in *DDX3X* may contribute to WNT-subgroup medulloblastoma by increasing LRLP proliferation rather than perturbing the migration of their daughter cells. Notably, comparable knockdown *in utero* of *Mll2*, *Gabrg1* and *Kdm6a* that were selectively mutated in non-WNT medulloblastomas had no apparent impact on LRLPs; supporting the value of our assay for assessing WNT-subgroup specific mutations and underscoring the importance of cell context for functional studies of genes mutated in cancer subgroups.

PIK3CA mutations promote WNT-subgroup medulloblastoma

Cancer-associated, activating mutations in *PIK3CA* were detected in a single case each of WNT-subgroup (PIK3CA(Q546K)), SHH-subgroup (PIK3CA(H1047R)) and subgroup-4 (PIK3CA(N345K)) medulloblastoma (Fig. 1 and Supplementary Fig. 23). Although *PIK3CA* mutations are common in adult cancers[40] and reported in medulloblastoma[41], their role in tumorigenesis remains controversial. In particular it is not known if these mutations initiate or progress cancer. To test this, we generated mice that express a conditional allele of the $Pik3ca^{E545K}$ mutation. Mice harbouring $Pik3ca^{E545K}$ or $Pik3ca^{E545K}$ and $Tp53^{flx/flx}$ were bred with *Blbp-Cre*, which drives efficient recombination in LRLPs[5]. *Blbp-Cre*;$Pik3ca^{E545K}$ mice, with or without $Tp53^{flx/flx}$, survived tumour free for a median of 212 days with no evidence of aberrant LRLP migration (Fig. 4a and data not shown). In stark contrast, 100% ($n = 11/11$) of *Blbp-Cre*;*Ctnnb1*$^{+/lox(Ex3)}$;$Tp53^{+/flx}$;$Pik3ca^{E545K}$ mice developed WNT-subgroup medulloblastomas by 3 months of age; only 4% ($n = 2/54$) of *Blbp-Cre*;*Ctnnb1*$^{+/lox(Ex3)}$;$Tp53^{+/flx}$ mice develop WNT-subgroup medulloblastoma by 11 months (Fig. 4a, b). *Pik3ca* wild-type and mutant mouse medulloblastomas displayed similar 'classic' histologies and nuclear Ctnnb1^{+}, but $Pik3ca^{E545K}$ mutant tumours contained greater AKT pathway activity as measured by pS6 and p4EBP1 immunostaining. Thus mutations in *PIK3CA* probably activate the AKT pathway to progress, rather than initiate, WNT-subgroup medulloblastoma.

SHH-subgroup medulloblastomas

Four of thirteen SHH-subgroup medulloblastomas contained expected biallelic inactivating alterations in *SUFU* or *PTCH1*. What drives aberrant SHH signals in the remaining cases remains unclear. These tumours contained mutations in *MLL2*, *TP53* and *PTEN* that have been reported previously in medulloblastoma[42]; but these mutations occur in other subgroups and are not known to activate SHH signals. Two SHH-subgroup tumours (nos 11 and 12) contained identical novel T48M mutations in the $GABA_A$ (γ-aminobutyric acid, subtype A) receptor, γ1, which is predicted to be deleterious (Fig. 1 and Supplementary Table 9). Disruption of $GABA_A$ receptors can enhance neural stem cell proliferation[43], suggesting that these mutations might deregulate the proliferation of GNPs that generate SHH-subgroup medulloblastomas.

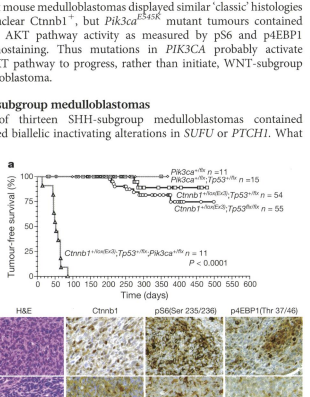


Figure 4 | *Pik3ca*^{E545K} accelerates but does not initiate WNT-subgroup medulloblastoma. **a**, Tumour-free survival of mice of the indicated genotype. All mice carry the *Blbp-cre* allele. Log rank $P < 0.0001$. **b**, Haematoxylin and eosin (H&E) and immunohistochemical stains of indicated tumours. Scale bar, 50 μm.

Discussion

We have identified several, new, recurrent, somatic mutations in specific subgroups of medulloblastoma. Alterations affecting *EZH2*, *KDM6A*, *CHD7* and *ZMYM3* seem to disrupt chromatin marking of genes in subgroup-3 and -4 tumours. Further epigenetic studies will be required to uncover the identity of these genes, but evidence suggests these may include *OTX2*, *MYC* and *MYCN*[44,45]. As amplification of these genes was detected almost exclusively in subgroup-3 and -4 tumours that lacked mutations in *KDM6A*, *CHD7* or *ZMYM3*, it is tempting to speculate that these genetic alterations target common transforming pathways. A recent study detected recurrent mutations in three other chromatin remodellers in medulloblastoma[42]: *SMARCA4*, *MLL2* and *MLL3*, but this study did not include details of tumour subgroup. Here, we show that mutations in *SMARCA4*, *CREBBP*, *TRRAP* and *MED13* are enriched in WNT-subgroup medulloblastomas; thereby uncovering potential cooperative mutations in chromatin remodellers and their binding-partner oncogene, CTNNB1. Thus, disruptions in the epigenetic machinery of medulloblastoma are likely to be subgroup specific and may cooperate with other oncogenic mutations. The low incidence of *MLL2* mutations detected in our study relative to previous work[42] probably reflects differences in study populations (see Supplementary Results).

Although medulloblastoma is more prevalent in males, especially with subgroup-3 and -4 disease[46], the reason for this sex bias is unknown. One potential explanation is the location of medulloblastoma oncogenes or tumour suppressor genes on chromosome X[47]. Three of the most recurrently mutated genes detected in our study are located on chromosome X, of which two (*ZMYM3* and *KDM6A*) were observed almost exclusively in males. Mutation in these genes might explain some of the male sex bias in medulloblastoma. The third mutated X chromosome gene, *DDX3X*, is more likely to be a WNT-subgroup medulloblastoma oncogene. Three of four female medulloblastomas carried heterozygous mutations in *DDX3X* that escape X inactivation[25], and our functional data indicate that mutations in this gene provide a proliferative advantage to LRLPs that generate these tumours.

Our findings also have important implications for drug development. Inhibitors of the epigenetic machinery, especially those that maintain H3K27me3—for example, EZH2 methylase—may be useful treatments for subgroup-3 and -4 disease. These tumours include the most aggressive forms of medulloblastoma, for which treatment options are limited. Mutations that activate PIK3CA and DDX3X in WNT-subgroup tumours might also be targeted with novel therapeutic strategies[48,49]. Future clinical trials of drugs that target these mutant proteins must recruit the appropriate patient populations, as we demonstrate that mutations show subgroup specificity in medulloblastoma. Our accurate mouse models of WNT-subgroup, SHH-subgroup and subgroup-3 medulloblastoma should help with future studies of the biological and therapeutic importance of the novel genetic alterations described in this study.

METHODS SUMMARY

Human tumour and matched blood samples were obtained with informed consent through an institutional review board approved protocol at St Jude Children's Research Hospital. WGS and analysis of WGS data were performed

as previously described[26]. Details of sequence coverage, custom capture and other validation procedures are provided in Supplementary Information (Supplementary Tables 12–15). Immunohistochemistry and immunofluorescence of human and mouse tissues were performed using routine techniques and primary antibodies of the appropriate tissues as described (Supplementary Methods). Medulloblastoma mRNA and DNA profiles were generated using Affymetrix U133v2 and SNP 6.0 arrays, respectively (Supplementary Methods). Real-time PCR with reverse transcriptase (RT–PCR) analysis of genes targeted in mouse LRLPs by shRNAs were performed as described previously[32]. LRLPs were isolated and transduced with indicated lentiviruses in stem cell cultures or targeted *in utero* with shRNAs or mutant cDNA sequences by electroporation as described[5] (Supplementary Information). Mice harbouring a Cre-inducible *Ptk3ca*^{E545K} allele were generated using homologous recombination: a lox-puro-STOP-lox cassette was introduced immediately upstream of the exon containing the initiation codon, exon 9 was replaced with an exon containing the E545K mutation. *Ptk3ca*^{E545K} mice were bred with *Blbp-Cre*;*Ctnnb1*$^{lox(Ex3)/lox(Ex3)}$ and *Tp53*$^{flx/flx}$ mice to generate progeny of the appropriate genotype and subjected to clinical surveillance.

Received 13 January; accepted 2 May 2012.
Published online 20 June 2012.

1. Central Brain Tumor Registry of the United States. Statistical report: primary brain tumors in the United States, 1995–1999. https://http://www.cbtrus.org/reports/2002/2002report.pdf (CBTRUS, 2006).
2. Taylor, M. D. *et al.* Molecular subgroups of medulloblastoma: the current consensus. *Acta Neuropathol.* **123,** 465–472 (2012).
3. Schüller, U. *et al.* Acquisition of granule neuron precursor identity is a critical determinant of progenitor cell competence to form Shh-induced medulloblastoma. *Cancer Cell* **14,** 123–134 (2008).
4. Yang, Z. J. *et al.* Medulloblastoma can be initiated by deletion of Patched in lineage-restricted progenitors or stem cells. *Cancer Cell* **14,** 135–145 (2008).
5. Gibson, P. *et al.* Subtypes of medulloblastoma have distinct developmental origins. *Nature* **468,** 1095–1099 (2010).
6. Kawauchi, D. *et al.* A mouse model of the most aggressive subgroup of human medulloblastoma. *Cancer Cell* **21,** 168–180 (2012).
7. Mulhern, R. K. *et al.* Neurocognitive consequences of risk-adapted therapy for childhood medulloblastoma. *J. Clin. Oncol.* **23,** 5511–5519 (2005).
8. Wang, J. *et al.* CREST maps somatic structural variation in cancer genomes with base-pair resolution. *Nature Methods* **8,** 652–654 (2011).
9. Rausch, T. *et al.* Genome sequencing of pediatric medulloblastoma links catastrophic DNA rearrangements with *TP53* mutations. *Cell* **148,** 59–71 (2012).
10. Castellino, R. C. *et al.* Heterozygosity for *Pten* promotes tumorigenesis in a mouse model of medulloblastoma. *PLoS ONE* **5,** e10849 (2010).
11. Hahn, H. *et al.* Mutations of the human homolog of *Drosophila* patched in the nevoid basal cell carcinoma syndrome. *Cell* **85,** 841–851 (1996).
12. Malkin, D. *et al.* Germ line p53 mutations in a familial syndrome of breast cancer, sarcomas, and other neoplasms. *Science* **250,** 1233–1238 (1990).
13. Hamilton, S. R. *et al.* The molecular basis of Turcot's syndrome. *N. Engl. J. Med.* **332,** 839–847 (1995).
14. Taylor, M. D. *et al.* Medulloblastoma in a child with Rubenstein-Taybi syndrome: case report and review of the literature. *Pediatr. Neurosurg.* **35,** 235–238 (2001).
15. Mikkelsen, T. S. *et al.* Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448,** 553–560 (2007).
16. Cao, R. *et al.* Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science* **298,** 1039–1043 (2002).
17. Czermin, B. *et al. Drosophila* Enhancer of Zeste/ESC complexes have a histone H3 methyltransferase activity that marks chromosomal Polycomb sites. *Cell* **111,** 185–196 (2002).
18. Agger, K. *et al.* UTX and JMJD3 are histone H3K27 demethylases involved in *HOX* gene regulation and development. *Nature* **449,** 731–734 (2007).
19. Schnetz, M. P. *et al.* Genomic distribution of CHD7 on chromatin tracks H3K4 methylation patterns. *Genome Res.* **19,** 590–601 (2009).
20. Sauvageau, M. & Sauvageau, G. Polycomb group proteins: multi-faceted regulators of somatic stem cells and cancer. *Cell Stem Cell* **7,** 299–313 (2010).
21. Morin, R. D. *et al.* Somatic mutations altering EZH2 (Tyr641) in follicular and diffuse large B-cell lymphomas of germinal-center origin. *Nature Genet.* **42,** 181–185 (2010).
22. Kleer, C. G. *et al.* EZH2 is a marker of aggressive breast cancer and promotes neoplastic transformation of breast epithelial cells. *Proc. Natl Acad. Sci. USA* **100,** 11606–11611 (2003).
23. Varambally, S. *et al.* The polycomb group protein EZH2 is involved in progression of prostate cancer. *Nature* **419,** 624–629 (2002).
24. van Haaften, G. *et al.* Somatic mutations of the histone H3K27 demethylase gene *UTX* in human cancer. *Nature Genet.* **41,** 521–523 (2009).
25. Yang, F., Babak, T., Shendure, J. & Disteche, C. M. Global survey of escape from X inactivation by RNA-sequencing in mouse. *Genome Res.* **20,** 614–622 (2010).
26. Christensen, J. *et al.* RBP2 belongs to a family of demethylases, specific for tri- and dimethylated lysine 4 on histone 3. *Cell* **128,** 1063–1076 (2007).
27. Lee, M. G., Wynder, C., Cooch, N. & Shiekhattar, R. An essential role for CoREST in nucleosomal histone 3 lysine 4 demethylation. *Nature* **437,** 432–435 (2005).
28. Mosimann, C., Hausmann, G. & Basler, K. β-Catenin hits chromatin: regulation of Wnt target gene activation. *Nature Rev. Mol. Cell Biol.* **10,** 276–286 (2009).
29. Hecht, A., Vleminckx, K., Stemmler, M. P., van Roy, F. & Kemler, R. The p300/CBP acetyltransferases function as transcriptional coactivators of β-catenin in vertebrates. *EMBO J.* **19,** 1839–1850 (2000).
30. Barker, N. *et al.* The chromatin remodelling factor Brg-1 interacts with β-catenin to promote target gene activation. *EMBO J.* **20,** 4935–4943 (2001).
31. Carrera, I., Janody, F., Leeds, N., Duveau, F. & Treisman, J. E. Pygopus activates Wingless target gene transcription through the mediator complex subunits Med12 and Med13. *Proc. Natl Acad. Sci. USA* **105,** 6644–6649 (2008).
32. Thompson, M. C. *et al.* Genomics identifies medulloblastoma subgroups that are enriched for specific genetic alterations. *J. Clin. Oncol.* **24,** 1924–1931 (2006).
33. Kool, M. *et al.* Integrated genomics identifies five medulloblastoma subtypes with distinct genetic profiles, pathway signatures and clinicopathological features. *PLoS ONE* **3,** e3088 (2008).
34. Orsulic, S., Huber, O., Aberle, H., Arnold, S. & Kemler, R. E-cadherin binding prevents β-catenin nuclear localization and β-catenin/LEF-1-mediated transactivation. *J. Cell Sci.* **112,** 1237–1245 (1999).
35. Risinger, J. I., Berchuck, A., Kohler, M. F. & Boyd, J. Mutations of the E-cadherin gene in human gynecologic cancers. *Nature Genet.* **7,** 98–102 (1994).
36. Becker, K.-F. *et al.* E-Cadherin gene mutations provide clues to diffuse type gastric carcinomas. *Cancer Res.* **54,** 3845–3852 (1994).
37. Pek, J. W. & Kai, T. DEAD-box RNA helicase Belle/DDX3 and the RNA interference pathway promote mitotic chromosome segregation. *Proc. Natl Acad. Sci. USA* **108,** 12007–12012 (2011).
38. Lai, M. C., Chang, W. C., Shieh, S. Y. & Tarn, W. Y. DDX3 regulates cell growth through translational control of cyclin E1. *Mol. Cell. Biol.* **30,** 5444–5453 (2010).
39. Schröder, M. Human DEAD-box protein 3 has multiple functions in gene regulation and cell cycle control and is a prime target for viral manipulation. *Biochem. Pharmacol.* **79,** 297–306 (2010).
40. Samuels, Y. *et al.* High frequency of mutations of the *PIK3CA* gene in human cancers. *Science* **304,** 554 (2004).
41. Broderick, D. K. *et al.* Mutations of PIK3CA in anaplastic oligodendrogliomas, high-grade astrocytomas, and medulloblastomas. *Cancer Res.* **64,** 5048–5050 (2004).
42. Parsons, D. W. *et al.* The genetic landscape of the childhood cancer medulloblastoma. *Science* **331,** 435–439 (2011).
43. Andäng, M. *et al.* Histone H2AX-dependent $GABA_A$ receptor regulation of stem cell proliferation. *Nature* **451,** 460–464 (2008).
44. Pasini, D. *et al.* Coordinated regulation of transcriptional repression by the RBP2 H3K4 demethylase and Polycomb-Repressive Complex 2. *Genes Dev.* **22,** 1345–1355 (2008).
45. Khan, A., Shover, W. & Goodliffe, J. M. Su(z)2 antagonizes auto-repression of Myc in *Drosophila*, increasing Myc levels and subsequent trans-activation. *PLoS ONE* **4,** e5076 (2009).
46. Northcott, P. A. *et al.* Medulloblastoma comprises four distinct molecular variants. *J. Clin. Oncol.* **29,** 1408–1414 (2011).
47. Speicher, M. R. & Carter, N. P. The new cytogenetics: blurring the boundaries with molecular biology. *Nature Rev. Cancer* **4,** 617–629 (2004).
48. Lindqvist, L. *et al.* Selective pharmacological targeting of a DEAD box RNA helicase. *PLoS One* **3,** e1583 (2008).
49. Engelman, J. A. Targeting PI3K signalling in cancer: opportunities, challenges and limitations. *Nature Rev. Cancer* **9,** 550–562 (2009).
50. Zhang, J. *et al.* The genetic basis of early T-cell precursor acute lymphoblastic leukaemia. *Nature* **481,** 157–163 (2012).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This research was supported as part of the St Jude Children's Research Hospital, Washington University Pediatric Cancer Genome Project. This work was supported by grants from the National Institutes of Health (R01CA129541, P01CA96832 and P30CA021765; R.J.G.), the Collaborative Ependymoma Research Network (CERN), Musicians against Childhood Cancer (MACC), The Noyes Brain Tumour Foundation, and by the American Lebanese Syrian Associated Charities (ALSAC). We are grateful to S. Temple for the gift of reagents and the staff of the Hartwell Center for Bioinformatics and Biotechnology and ARC at St Jude Children's Research Hospital for technical assistance.

Author Contributions G.R., M.P., T.A.K., C.L., X.C., L.D., T.N.P., E.H., L.W., X.Z., N.Cb., R.H., N.Cu., R.T., J.W., G.W., M.R., X.H., J.B., P.G., J.M., J.E., B.V., A.O.-T., T.L., S.Po., S.Pa., D.Z., D.K. and D.F. contributed to the design and conduct of experiments and to the writing. S.J.B., R.K., M.F.R., R.S.F., L.L.F., D.J.D., K.O. and E.R.M. contributed to experimental design and to the writing. A.G., D.W.E., C.C.L., E.B., T.H., S.G. and R.C. provided clinical expertise. R.K.W., J.R.D., J.Z. and R.J.G. conceived the research and contributed to the design, direction and reporting of the study.

Author Information Sequence and SNP array data were deposited in dbGaP under accession number phs000409 and in the Sequence Read Archive (SRA) under accession number SRP008292. Reprints and permissions information is available at www.nature.com/reprints. This paper is distributed under the terms of the Creative Commons Attributions-Non-Commercial-Share Alike licence, and is freely available to all readers at www.nature.com/nature. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to R.J.G. (Richard.Gilbertson@stjude.org) or J.Z. (Jinghui.Zhang@stjude.org).

ARTICLE

doi:10.1038/nature11327

Subgroup-specific structural variation across 1,000 medulloblastoma genomes

A list of authors and their affiliations appears at the end of the paper

Medulloblastoma, the most common malignant paediatric brain tumour, is currently treated with nonspecific cytotoxic therapies including surgery, whole-brain radiation, and aggressive chemotherapy. As medulloblastoma exhibits marked intertumoural heterogeneity, with at least four distinct molecular variants, previous attempts to identify targets for therapy have been underpowered because of small samples sizes. Here we report somatic copy number aberrations (SCNAs) in 1,087 unique medulloblastomas. SCNAs are common in medulloblastoma, and are predominantly subgroup-enriched. The most common region of focal copy number gain is a tandem duplication of *SNCAIP*, a gene associated with Parkinson's disease, which is exquisitely restricted to Group 4*α*. Recurrent translocations of *PVT1*, including *PVT1-MYC* and *PVT1-NDRG1*, that arise through chromothripsis are restricted to Group 3. Numerous targetable SCNAs, including recurrent events targeting TGF-β signalling in Group 3, and NF-κB signalling in Group 4, suggest future avenues for rational, targeted therapy.

Brain tumours are the most common cause of childhood oncological death, and medulloblastoma is the most common malignant paediatric brain tumour. Current medulloblastoma therapy including surgical resection, whole-brain and spinal cord radiation, and aggressive chemotherapy supplemented by bone marrow transplant yields five-year survival rates of 60–70%[1]. Survivors are often left with significant neurological, intellectual and physical disabilities secondary to the effects of these nonspecific cytotoxic therapies on the developing brain[2].

Recent evidence suggests that medulloblastoma actually comprises multiple molecularly distinct entities whose clinical and genetic differences may require separate therapeutic strategies[3–6]. Four principal subgroups of medulloblastoma have been identified: WNT, SHH, Group 3 and Group 4 (ref. 7), and there is preliminary evidence for clinically significant subdivisions of the subgroups[3,7,8]. Rational, targeted therapies based on genetics are not currently in use for medulloblastoma, although inhibitors of the Sonic Hedgehog pathway protein Smoothened have shown early promise[9]. Actionable targets for WNT, Group 3 and Group 4 tumours have not been identified[4,10]. Sanger sequencing of 22 medulloblastoma exomes revealed on average only 8 single nucleotide variants (SNVs) per tumour[11]. Some SNVs were subgroup-restricted (*PTCH1*, *CTNNB1*), whereas others occurred across subgroups (*TP53*, *MLL2*). We proposed that the observed intertumoural heterogeneity might have underpowered prior attempts to discover targets for rational therapy.

The Medulloblastoma Advanced Genomics International Consortium (MAGIC) consisting of scientists and physicians from 46 cities across the globe gathered more than 1,200 medulloblastomas which were studied by SNP arrays ($n = 1{,}239$; Fig. 1a, Supplementary Fig. 1 and Supplementary Tables 1–3). Medulloblastoma subgroup affiliation of 827 cases was determined using a custom nanoString-based RNA assay (Supplementary Fig. 2)[12]. Disparate patterns of broad cytogenetic gain and loss were observed across the subgroups (Fig. 1b and Supplementary Figs 3, 7, 8, 10 and 11). Analysis of the entire cohort using GISTIC2 (ref. 13) to discover significant 'driver' events delineated 62 regions of recurrent SCNA (Fig. 1c, Supplementary Fig. 4 and Supplementary Tables 4 and 5); analysis by subgroup increased sensitivity such that 110 candidate 'driver' SCNAs were identified, most of which are subgroup-enriched (Fig. 1c–e and Supplementary Table 6).

Twenty-eight regions of recurrent high-level amplification (copy number ≥ 5) were identified (Fig. 1d and Supplementary Table 7). The most prevalent amplifications affected members of the MYC family with *MYCN* predominantly amplified in SHH and Group 4, *MYC* in Group 3, and *MYCL1* in SHH medulloblastomas. Multiple genes/regions were exclusively amplified in SHH, including *GLI2*, *MYCL1*, *PPM1D*, *YAP1* and *MDM4* (Fig. 1d). Recurrent homozygous deletions were exceedingly rare, with only 15 detected across 1,087 tumours (Fig. 1e). Homozygous deletions targeting known tumour suppressors *PTEN*, *PTCH1* and *CDKN2A/B* were the most common, all enriched in SHH cases (Fig. 1e and Supplementary Table 7). Novel homozygous deletions included *KDM6A*, a histone-lysine demethylase deleted in Group 4. A custom nanoString CodeSet was used to verify 24 significant regions of gain across 192 MAGIC cases, resulting in a verification rate of 90.9% (Supplementary Fig. 5). We conclude that SCNAs in medulloblastoma are common, and are predominantly subgroup-enriched.

Subgroup-specific SCNAs in medulloblastoma

WNT medulloblastoma genomes are impoverished of recurrent focal regions of SCNA, exhibiting no significant regions of deletion and only a small subset of focal gains found at comparable frequencies in non-WNT tumours (Supplementary Figs 4, 6 and Supplementary Table 8). *CTNNB1* mutational screening confirmed canonical exon 3 mutations in 63 out of 71 (88.7%) WNT tumours, whereas monosomy 6 was detected in 58 out of 76 (76.3%) (Supplementary Fig. 6; Supplementary Table 9). Four WNT tumours (4/71; 5.6%) had neither *CTNNB1* mutation nor monosomy 6, but maintained typical WNT expression signatures. Given the size of our cohort and the resolution of the platform, we conclude that there are no frequent, targetable SCNAs for WNT medulloblastoma.

SHH tumours exhibit multiple significant focal SCNAs (Fig. 2a, Supplementary Figs 12, 15, 16 and Supplementary Tables 10 and 11). SHH enriched/restricted SCNAs included amplification of *GLI2* and deletion of *PTCH1* (Fig. 2a, e, f)[10]. *MYCN* and *CCND2* were among the most frequently amplified genes in SHH (Supplementary Table 6), but were also altered in non-SHH cases. Genes upregulated in SHH tumours (that is, SHH signature genes) are significantly over-represented among the genes focally amplified in SHH tumours

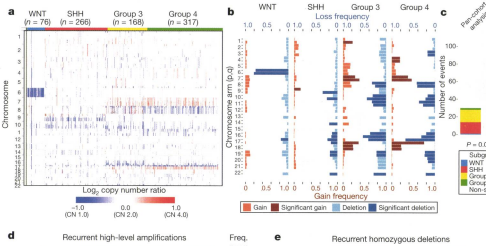


Figure 1 | Genomic heterogeneity of medulloblastoma subgroups. **a**, The medulloblastoma genome classified by subgroup. **b**, Frequency and significance (Q value ≤ 0.1) of broad cytogenetic events across medulloblastoma subgroups. **c**, Significant regions of focal SCNA identified by GISTIC2 in either pan-cohort or subgroup-specific analyses. **d**, **e**, Recurrent high-level amplifications (**d**, segmented copy number (CN) ≥ 5) and homozygous deletions (**e**, segmented CN ≤ 0.7) in medulloblastoma. The number of genes mapping to the GISTIC2 peak region (where applicable) is listed in brackets after the suspected driver gene, as is the frequency of each event.

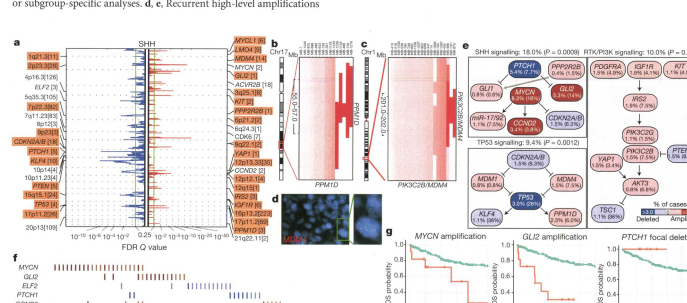


Figure 2 | Genomic alterations affect core signalling pathways in SHH medulloblastoma. **a**, GISTIC2 significance plot of amplifications (red) and deletions (blue) observed in SHH. The number of genes mapping to each significant region are included in brackets and regions enriched in SHH are shaded red. **b**, **c**, Recurrent amplifications of *PPM1D* (**b**) and *PIK3C2B*/*MDM4* (**c**) are restricted to SHH. **d**, Fluorescence *in situ* hybridization (FISH) validation of *MDM4* amplification. **e**, SHH signalling, TP53 signalling and RTK/PI3K signalling represent the core pathways genomically targeted in SHH. *P* values indicate the prevalence with which the respective pathway is targeted in SHH versus non-SHH cases (Fisher's exact test). Frequencies of focal and broad (parentheses) SCNAs are listed. **f**, Mutual exclusivity analysis of focal SCNAs in SHH. **g**, Clinical implications of SCNAs affecting *MYCN*, *GLI2* or *PTCH1* in SHH (log-rank tests).

($P = 0.001–0.02$, permutation tests; Supplementary Fig. 9). Recurrent amplification of SHH signature genes has clinical implications, as amplification of downstream transcriptional targets could mediate resistance to upstream SHH pathway inhibitors[14].

Novel, SHH-enriched SCNAs included components of TP53 signalling, including amplifications of *MDM4* and *PPM1D*, and focal deletions of *TP53* (Fig. 2a–e). Targetable events, including amplifications of IGF signalling genes *IGF1R* and *IRS2*, PI3K genes *PIK3C2G* and *PIK3C2B*, and deletion of *PTEN* were restricted to SHH tumours (Fig. 2a, c, e). Importantly, focal events affecting genes in the SHH pathway were largely mutually exclusive and prognostically significant (Fig. 2f, g). Many of the recurrent, targetable SCNAs identified in SHH medulloblastoma (*IGF1R*, *KIT*, *MDM4*, *PDGFRA*, *PIK3C2G*, *PIK2C2B* and *PTEN*) have already been targeted with small molecules for treatment of other malignancies, which might allow rapid translation for targeted therapy of subsets of SHH patients (Supplementary Table 16). Novel SHH targets identified here are excellent candidates for combinatorial therapy with Smoothened inhibitors, to avoid the resistance encountered in both humans and mice[8,14,15].

Group 3 and Group 4 medulloblastomas have generic names as comparatively little is known about their genetic basis, and no targets for rational therapy have been identified[7]. *MYC* amplicons are largely restricted to Group 3, whereas *MYCN* amplicons are seen in Group 4 and SHH tumours (Fig. 1d)[3,4]. Indeed, *MYC* and *MYCN* loci comprise the most significant regions of amplification observed in Group 3 and Group 4, respectively (Fig. 3a, b, Supplementary Figs 13, 14, 17–20 and Supplementary Tables 12–15). Group 3 *MYC* amplicons were mutually exclusive from those affecting the known medulloblastoma oncogene *OTX2* (ref. 16) and were highly prognostic (Supplementary Fig. 21)[3,16]. Type II activin receptors, *ACVR2A* and *ACVR2B* and family member *TGFBR1* are highly amplified in Group 3 tumours, indicating deregulation of TGF-β signalling as a driver event in Group 3 (Fig. 3c–e and Supplementary Fig. 22). The Group 3-enriched medulloblastoma oncogene *OTX2* is a prominent target of TGF-β signalling in the developing nervous system[17] and TGF-β pathway inhibitors *CD109* (ref. 18), *FKBP1A* (refs 19 and 20) and *SNX6* (ref. 20) are recurrently deleted in Group 3 (Fig. 3a, d). SCNAs in TGF-β pathway genes were heavily enriched in Group 3 ($P = 5.37 \times 10^{-5}$, Fisher's exact test) and found in at least 20.2% of cases, indicating that TGF-β signalling represents the first rational target for this poor prognosis subgroup (Fig. 3d). Similarly, novel deletions affecting regulators of the NF-κB pathway, including *NFKBIA* (ref. 21) and *USP4* (ref. 22) were identified in Group 4 (Supplementary Fig. 23), proposing that NF-κB signalling may represent a rational Group 4 therapeutic target.

Network analysis of Group 3 and Group 4 SCNAs illustrates the different pathways over-represented in each subgroup. Only TGF-β signalling is unique to Group 3 (Fig. 3e). In contrast, cell-cycle control, chromatin modification and neuronal development are all Group 4-enriched. Cumulatively, the dismal prognosis of Group 3 patients, the lack of published targets for rational therapy, and the prior targeting of TGF-β signalling in other diseases suggest that TGF-β may represent an appealing target for Group 3 rational therapies (Supplementary Table 16).

SNCAIP tandem duplication is common in Group 4

Although Group 4 is the most prevalent medulloblastoma subgroup, its pathogenesis remains poorly understood. The most frequent SCNA observed in Group 4 (33/317; 10.4%) is a recurrent region of single copy gain on chr5q23.2 targeting a single gene, *SNCAIP* (synuclein, alpha interacting protein) (Fig. 4a and Supplementary Fig. 24). *SNCAIP*, encodes synphilin-1, which binds to α-synuclein to promote the formation of Lewy bodies in the brains of patients with Parkinson's disease[23,24]. Additionally, rare germline mutations of *SNCAIP* have been described in Parkinson's families[25]. Large insert, mate-pair, whole-genome sequencing (WGS) demonstrates that

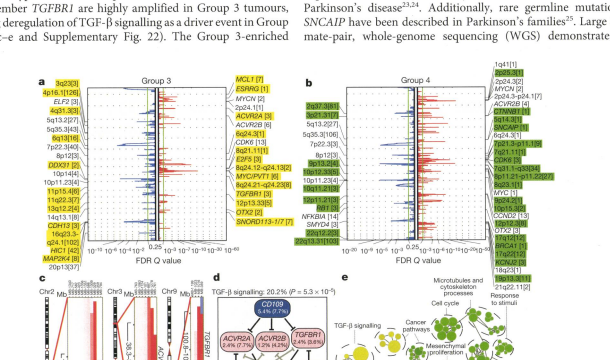


Figure 3 | The genomic landscape of Group 3 and Group 4 medulloblastoma. a, **b**, GISTIC2 plots depicting significant SCNAs in Group 3 (**a**) and Group 4 (**b**) with subgroup-enriched regions shaded in yellow and green, respectively. **c**, Recurrent amplifications targeting type II (*ACVR2A* and *ACVR2B*) and type I (*TGFBR1*) activin receptors in Group 3. **d**, Recurrent SCNAs affecting the TGF-β pathway in Group 3 ($P = 5.73 \times 10^{-5}$, Fisher's exact test). Frequencies of focal and broad (parentheses) SCNAs are listed. **e**, Enrichment map of gene sets affected by SCNAs in Group 3 versus Group 4.

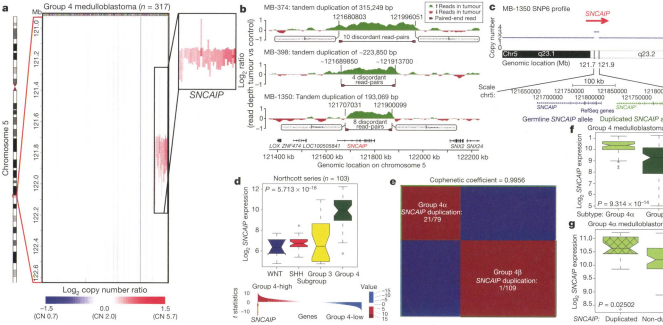


Figure 4 | Tandem duplication of *SNCAIP* defines a novel subtype of Group 4. **a**, Highly recurrent, focal, single-copy gain of *SNCAIP* in Group 4. **b**, Paired-end mapping verifies recurrent tandem duplication of *SNCAIP* in Group 4. **c**, Schematic representation of *SNCAIP* tandem duplication. **d**, *SNCAIP* is a Group 4 signature gene. Upper panel, *SNCAIP* expression across subgroups in a published series of 103 primary medulloblastomas. Error bars depict the minimum and maximum values, excluding outliers. Lower panel, *SNCAIP* ranks among the top 1% (rank, 39th out of 16,758) of highly expressed genes in Group 4. **e**, NMF consensus clustering of 188 expression-profiled Group 4 tumours supports two transcriptionally distinct subtypes designated 4α and 4β (cophenetic coefficient = 0.9956). 21 out of 22 *SNCAIP* duplicated cases belong to Group 4α ($P = 3.12 \times 10^{-8}$, Fisher's exact test). **f**, *SNCAIP* expression is significantly elevated in Group 4α versus 4β ($P = 9.31 \times 10^{-14}$, Mann–Whitney test). **g**, Group 4α cases harbouring *SNCAIP* duplication exhibit a ~1.5-fold increase in *SNCAIP* expression. **f**, **g**, Error bars depict the minimum and maximum values, excluding outliers.

SNCAIP copy number gains arise from tandem duplication of a truncated *SNCAIP* (lacking non-coding exon 1), inserted telomeric to the germline *SNCAIP* allele (Fig. 4b, c and Supplementary Fig. 25). Affymetrix SNP6 array profiling of patient-matched germline material confirmed that *SNCAIP* duplications are somatic (Supplementary Fig. 26), and subsequent whole-transcriptome sequencing (RNA-Seq) of select Group 4 cases ($n = 5$) verified that *SNCAIP* is the only gene expressed in the duplicated region (Supplementary Fig. 27). Analysis of published copy number profiles for 3,131 primary tumours[26] and 947 cancer cell lines[27] (total of 4,078 cases) revealed only four cases with apparent duplication of *SNCAIP*, all of which were inferred as Group 4 medulloblastomas (data not shown). We conclude that *SNCAIP* duplication is a somatic event highly specific to Group 4 medulloblastoma.

Re-analysis of 499 published medulloblastoma expression profiles confirmed that *SNCAIP* is one of the most highly upregulated Group 4 signature genes (Fig. 4d and Supplementary Fig. 28). Profiling of 188 Group 4 tumours on expression microarrays followed by consensus non-negative matrix factorization (NMF) clustering delineates two subtypes of Group 4 (4α and 4β; Fig. 4e and Supplementary Fig. 29). Strikingly, 21 out of 22 *SNCAIP* duplicated cases belonged to Group 4α ($P = 3.12 \times 10^{-8}$, Fisher's exact test). *SNCAIP* is more highly expressed in Group 4α than 4β (Fig. 4f), and 4α samples with tandem duplication showed approximately 1.5-fold increased expression, consistent with gene dosage (Fig. 4g and Supplementary Figs 35 and 36). Group 4α exhibits a relatively balanced genome compared to 4β (Supplementary Figs 30–32), and several 4α cases harbour *SNCAIP* duplication in conjunction with i17q and no other SCNAs (Supplementary Fig. 33). Importantly, *SNCAIP* duplications are mutually exclusive from other prominent SCNAs in Group 4, including *MYCN* and *CDK6* amplifications (Supplementary Fig. 34).

PVT1 fusions arise via chromothripsis in Group 3

Although recurrent gene fusions have recently been discovered in solid tumours, none have been reported in medulloblastoma. RNA-Seq of Group 3 tumours ($n = 13$) identified two independent gene fusions in two different tumours (MB-182 and MB-586), both involving the 5′ end of *PVT1*, a non-coding gene frequently co-amplified with *MYC* in Group 3 (Fig. 5a, b, Supplementary Fig. 37 and Supplementary Tables 17 and 18). Sanger sequencing confirmed a fusion transcript consisting of exons 1 and 3 of *PVT1* fused to the coding sequence of *MYC* (exons 2 and 3) in MB-182, and a fusion involving *PVT1* exon 1 fused to the 3′ end of *NDRG1* in MB-586 (Fig. 5a, b).

Group 3 copy number data at the *MYC*/*PVT1* locus indicated that additional samples might harbour *PVT1* gene fusions (Fig. 5c). PCR with reverse transcription (RT–PCR) profiling of select Group 3 cases confirmed *PVT1*-*MYC* fusions in at least 60% (12/20) of *MYC*-amplified cases (Fig. 5d and Supplementary Table 19). Fusion transcripts included many other portions of chr8q, with up to four different genomic loci mapping to a single transcript, a pattern reminiscent of chromothripsis[28,29] (Fig. 5d). WGS performed on four *MYC*-amplified Group 3 tumours harbouring *PVT1* fusion transcripts identified a series of complex genomic rearrangements on chr8q (Fig. 5e, f, Supplementary Fig. 38 and Supplementary Tables 20 and 21). Chromosome 8 copy number profile for MB-586 (*PVT1*-*NDRG1*) derived from WGS showed that *PVT1* and *NDRG1* are structurally linked, as predicted by RNA-Seq, and several adjacent regions of 8q24 were extensively rearranged (Fig. 5e, f and Supplementary Table 21). Monte Carlo simulation suggests that this fragmented 8q amplicon arose through chromothripsis, a process of erroneous DNA repair following a single catastrophic event in which a chromosome is shattered into many pieces (Supplementary Fig. 39). Further examination of our copy number data set revealed rare examples of chromothripsis across subgroups (Supplementary Fig. 40), with only chr8 in Group 3 demonstrating statistically significant, region-specific chromothripsis ($Q = 0.0004$, false discovery rate (FDR)-corrected Fisher's exact test). Among Group 3 tumours, the occurrence of chr8q chromothripsis is correlated with deletion of chr17p (location of *TP53*; data not shown), in keeping with the association of loss of

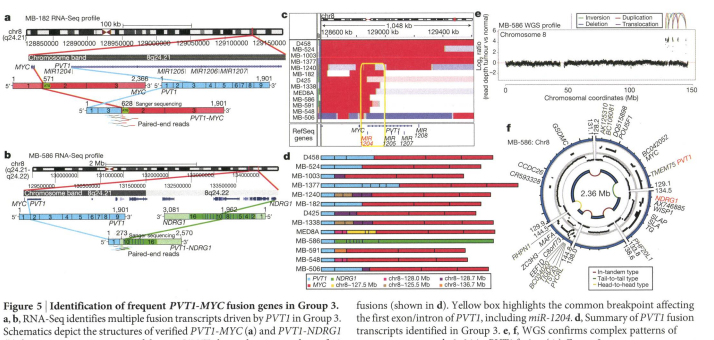


Figure 5 | Identification of frequent *PVT1-MYC* fusion genes in Group 3. **a**, **b**, RNA-Seq identifies multiple fusion transcripts driven by *PVT1* in Group 3. Schematics depict the structures of verified *PVT1-MYC* (**a**) and *PVT1-NDRG1* (**b**) fusion genes. **c**, Heat map of the *MYC*/*PVT1* locus showing a subset of 13 *MYC*-amplified Group 3 cases subsequently verified to exhibit *PVT1* gene fusions (shown in **d**). Yellow box highlights the common breakpoint affecting the first exon/intron of *PVT1*, including *miR-1204*. **d**, Summary of *PVT1* fusion transcripts identified in Group 3. **e**, **f**, WGS confirms complex patterns of rearrangement on chr8q24 in *PVT1* fusion (+) Group 3.

TP53 and chromothripsis recently described in medulloblastoma ($P = 0.0199$, Fisher's exact test)[28]. Whereas the *PVT1* locus has been suggested to be a genomically fragile site, we observe that the majority of *MYC*-amplified Group 3 tumours harbour *PVT1* fusions that arise through a process consistent with chromothripsis.

PVT1 is a non-coding host gene for four microRNAs, *miR-1204–miR-1207*. Previous studies have implicated *miR-1204* as a candidate oncogene that enhances oncogenesis in combination with *MYC*[30,31]. *PVT1* fusions identified in this study involve only *PVT1* exon 1 and *miR-1204*. Importantly, *miR-1204*, but not the adjacent *miR-1205* and *miR-1206*, is expressed at a higher level in *PVT1-MYC* fusion (+) Group 3 tumours compared to fusion (−) cases ($P = 0.0008$, Mann–Whitney test; Fig. 6a). To evaluate whether aberrant expression of *miR-1204* contributes to the malignant phenotype, we inhibited *miR-1204* in MED8A cells, a Group 3 medulloblastoma cell line with a confirmed *PVT1-MYC* fusion (Fig. 5d). Antagomir-mediated RNA interference of *miR-1204* had a pronounced effect on MED8A growth (Fig. 6b). A comparable reduction in proliferative capacity was achieved with knockdown of *MYC*. Conversely, the medulloblastoma cell line ONS76 exhibits neither *MYC* amplification nor a detectable *PVT1-MYC* fusion gene, and knockdown of *miR-1204* had no effect in this line (Fig. 6c).

PVT1 has been reported previously in fusion transcripts with a number of partners[30,32,33]. The most prevalent form of the *PVT1-MYC* fusion in Group 3 tumours lacks the first, non-coding exon of *MYC*, similar to forms of *MYC* that have been described in Burkitt's lymphoma[34] (Fig. 5a, d). The *PVT1* promoter contains two non-canonical E-boxes and can be activated by *MYC*[31]. This indicates a positive feedback model where *MYC* can reinforce its own expression from the *PVT1* promoter in *PVT1-MYC* fusion (+) tumours. Indeed, knockdown of *MYC* alone in MED8A cells resulted in diminished expression of both *MYC* and *miR-1204*, suggesting *MYC* may positively regulate *PVT1* (that is, *miR-1204*) expression in medulloblastoma cells (Supplementary Fig. 41).

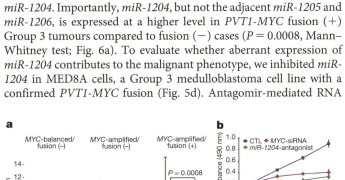


Figure 6 | Functional synergy between *miR-1204* and *MYC* secondary to *PVT1-MYC* fusion. **a**, Quantitative RT–PCR of *PVT1*-encoded microRNAs confirms upregulation of *miR-1204* in *PVT1-MYC* fusion (+) Group 3 tumours. *MYC*-balanced/fusion (−), $n = 4$; *MYC*-amplified/fusion (−), $n = 6$; *MYC*-amplified/fusion (+), $n = 8$. Error bars represent standard error of the mean (s.e.m.) and reflect variability among samples. **b**, **c**, Knockdown of *miR-1204* attenuates the proliferative capacity of *PVT1-MYC* fusion (+) MED8A medulloblastoma cells (**b**) but has no effect on fusion (−) ONS76 cells (**c**). Error bars represent the standard deviation (s.d.) of triplicate experiments. CTL, control.

Discussion

Medulloblastomas have few SNVs compared to many adult epithelial malignancies[11], whereas SCNAs seem to be quite common. Medulloblastoma is a heterogeneous disease[7], thereby requiring large cohorts to detect subgroup-specific events. Through the accumulation of >1,200 medulloblastomas in MAGIC, we have identified novel and significant SCNAs. Many of the significant SCNAs are subgroup-restricted, highly supporting their role as driver events in their respective subgroups.

Expression of synphilin-1 in neuronal cells results in decreased cell doubling time[35], decreased caspase-3 activation[36], decreased *TP53* transcriptional activity and messenger RNA levels, and decreased apoptosis[37]. Synphilin-1 is ubiquitinated by parkin, which is encoded by the hereditary Parkinson's disease gene *PARK2* (ref. 24), a candidate tumour suppressor gene[38]. Whereas patients with Parkinson's disease have an overall decreased risk of cancer, they may have an increased incidence of brain tumours[39,40]. As tandem duplications of *SNCAIP* are highly recurrent, stereotypical, subgroup-restricted, affect only a single gene, and as *SNCAIP*-duplicated tumours have few if any other SCNAs, *SNCAIP* is a probable driver gene, and merits investigation

as a target for therapy of Group 4α. Similarly, *PVT1* fusion genes are highly recurrent, restricted to Group 3, arise through a chromothripsis-like process, and are the first recurrent translocation reported in medulloblastoma.

We identify a number of highly targetable, recurrent, subgroup-specific SCNAs that could form the basis for future clinical trials (that is, PI3K signalling in SHH, TGF-β signalling in Group 3, and NF-κB signalling in Group 4). Activation of these pathways through alternative, currently unknown genetic and epigenetic events could increase the percentage of patients amenable to targeted therapy. We also identify a number of highly 'druggable' events that occur in a minority of cases. The cooperative, global approach of the MAGIC consortium has allowed us to overcome the barrier of intertumoural heterogeneity in an uncommon paediatric tumour, and to identify the relevant and targetable SCNAs for the affected children.

METHODS SUMMARY

All patient samples were obtained with consent as outlined by individual institutional review boards. Genomic DNA was prepared, processed and hybridized to Affymetrix SNP6 arrays according to manufacturer's instructions. Raw copy number estimates were obtained in dChip, followed by CBS segmentation in R. SCNAs were identified using GISTIC2 (ref. 13). Driver genes within SCNAs were inferred by integrating matched expressions, literature evidence and other data sets. Pathway enrichment of SCNAs was analysed with g:Profiler and visualized in Cytoscape using Enrichment Map. Fluorescence *in situ* hybridization (FISH) was performed as described previously[8,10]. Medulloblastoma subgroup was assigned using a custom nanoString CodeSet as described previously[22]. Tandem duplication of *SNCAIP* was confirmed by paired-end mapping as previously reported[38]. RNA was extracted, processed and hybridized to Affymetrix Gene 1.1 ST Arrays as recommended by the manufacturer. Consensus NMF clustering was performed in GenePattern. Gene fusions were identified from RNA-Seq data using Trans-ABySS. Medulloblastoma cell lines were maintained as described[16]. Proliferation assays were performed with the Promega CellTiter 96 Assay. Additional methods are detailed in full in Supplementary Methods.

Received 29 February; accepted 14 June 2012.
Published online 25 July 2012.

1. Gajjar, A. *et al.* Risk-adapted craniospinal radiotherapy followed by high-dose chemotherapy and stem-cell rescue in children with newly diagnosed medulloblastoma (St Jude Medulloblastoma-96): long-term results from a prospective, multicentre trial. *Lancet Oncol.* **7,** 813–820 (2006).
2. Mabbott, D. J. *et al.* Serial evaluation of academic and behavioral outcome after treatment with cranial radiation in childhood. *J. Clin. Oncol.* **23,** 2256–2263 (2005).
3. Cho, Y. J. *et al.* Integrative genomic analysis of medulloblastoma identifies a molecular subgroup that drives poor clinical outcome. *J. Clin. Oncol.* **29,** 1424–1430 (2011).
4. Northcott, P. A. *et al.* Medulloblastoma comprises four distinct molecular variants. *J. Clin. Oncol.* **29,** 1408–1414 (2011).
5. Remke, M. *et al. FSTL5* is a marker of poor prognosis in non-WNT/non-SHH medulloblastoma. *J. Clin. Oncol.* **29,** 3852–3861 (2011).
6. Northcott, P. A., Korshunov, A., Pfister, S. M. & Taylor, M. D. The clinical implications of medulloblastoma subgroups. *Nature Rev. Neurol.* **8,** 340–351 (2012).
7. Taylor, M. D. *et al.* Molecular subgroups of medulloblastoma: the current consensus. *Acta Neuropathol.* **123,** 465–472 (2011).
8. Northcott, P. A. *et al.* Pediatric and adult sonic hedgehog medulloblastomas are clinically and molecularly distinct. *Acta Neuropathol.* **122,** 231–240 (2011).
9. Rudin, C. M. *et al.* Treatment of medulloblastoma with hedgehog pathway inhibitor GDC-0449. *N. Engl. J. Med.* **361,** 1173–1178 (2009).
10. Northcott, P. A. *et al.* Multiple recurrent genetic events converge on control of histone lysine methylation in medulloblastoma. *Nature Genet.* **41,** 465–472 (2009).
11. Parsons, D. W. *et al.* The genetic landscape of the childhood cancer medulloblastoma. *Science* **331,** 435–439 (2011).
12. Northcott, P. A. *et al.* Rapid, reliable, and reproducible molecular sub-grouping of clinical medulloblastoma samples. *Acta Neuropathol.* **123,** 615–626 (2012).
13. Mermel, C. H. *et al.* GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol.* **12,** R41 (2011).
14. Buonamici, S. *et al.* Interfering with resistance to smoothened antagonists by inhibition of the PI3K pathway in medulloblastoma. *Sci. Transl. Med.* **2,** 51ra70 (2010).
15. Yauch, R. L. *et al.* Smoothened mutation confers resistance to a hedgehog pathway inhibitor in medulloblastoma. *Science* **326,** 572–574 (2009).
16. Adamson, D. C. *et al.* OTX2 is critical for the maintenance and progression of Shh-independent medulloblastomas. *Cancer Res.* **70,** 181–191 (2010).
17. Jia, S., Wu, D., Xing, C. & Meng, A. Smad2/3 activities are required for induction and patterning of the neuroectoderm in zebrafish. *Dev. Biol.* **333,** 273–284 (2009).
18. Bizet, A. A. *et al.* The TGF-β co-receptor, CD109, promotes internalization and degradation of TGF-β receptors. *Biochim. Biophys. Acta* **1813,** 742–753 (2011).
19. Wang, T., Donahoe, P. K. & Zervos, A. S. Specific interaction of type I receptors of the TGF-beta family with the immunophilin FKBP-12. *Science* **265,** 674–676 (1994).
20. Parks, W. T. *et al.* Sorting nexin 6, a novel SNX, interacts with the transforming growth factor-β family of receptor serine-threonine kinases. *J. Biol. Chem.* **276,** 19332–19339 (2001).
21. Bredel, M. *et al. NFKBIA* deletion in glioblastomas. *N. Engl. J. Med.* **364,** 627–637 (2011).
22. Xiao, N. *et al.* Ubiquitin-specific protease 4 (USP4) targets TRAF2 and TRAF6 for deubiquitination and inhibits TNFα-induced cancer cell migration. *Biochem. J.* **441,** 979–986 (2012).
23. Engelender, S. *et al.* Synphilin-1 associates with α-synuclein and promotes the formation of cytosolic inclusions. *Nature Genet.* **22,** 110–114 (1999).
24. Chung, K. K. *et al.* Parkin ubiquitinates the α-synuclein-interacting protein, synphilin-1: implications for Lewy-body formation in Parkinson disease. *Nature Med.* **7,** 1144–1150 (2001).
25. Marx, F. P. *et al.* Identification and functional characterization of a novel R621C mutation in the synphilin-1 gene in Parkinson's disease. *Hum. Mol. Genet.* **12,** 1223–1231 (2003).
26. Beroukhim, R. *et al.* The landscape of somatic copy-number alteration across human cancers. *Nature* **463,** 899–905 (2010).
27. Barretina, J. *et al.* The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* **483,** 603–607 (2012).
28. Rausch, T. *et al.* Genome sequencing of pediatric medulloblastoma links catastrophic DNA rearrangements with *TP53* mutations. *Cell* **148,** 59–71 (2012).
29. Stephens, P. J. *et al.* Massive genomic rearrangement acquired in a single catastrophic event during cancer development. *Cell* **144,** 27–40 (2011).
30. Shtivelman, E. & Bishop, J. M. The *PVT* gene frequently amplifies with *MYC* in tumor cells. *Mol. Cell. Biol.* **9,** 1148–1154 (1989).
31. Carramusa, L. *et al.* The PVT-1 oncogene is a Myc protein target that is overexpressed in transformed cells. *J. Cell. Physiol.* **213,** 511–518 (2007).
32. Shtivelman, E. & Bishop, J. M. Effects of translocations on transcription from *PVT*. *Mol. Cell. Biol.* **10,** 1835–1839 (1990).
33. Pleasance, E. D. *et al.* A small-cell lung cancer genome with complex signatures of tobacco exposure. *Nature* **463,** 184–190 (2010).
34. Hann, S. R., King, M. W., Bentley, D. L., Anderson, C. W. & Eisenman, R. N. A non-AUG translational initiation in *c-myc* exon 1 generates an N-terminally distinct protein whose synthesis is disrupted in Burkitt's lymphomas. *Cell* **52,** 185–195 (1988).
35. Li, X. *et al.* Synphilin-1 exhibits trophic and protective effects against Rotenone toxicity. *Neuroscience* **165,** 455–462 (2010).
36. Smith, W. W. *et al.* Synphilin-1 attenuates neuronal degeneration in the A53T alpha-synuclein transgenic mouse model. *Hum. Mol. Genet.* **19,** 2087–2098 (2010).
37. Gianne, E. *et al.* Caspase-3-derived C-terminal product of synphilin-1 displays antiapoptotic function via modulation of the p53-dependent cell death pathway. *J. Biol. Chem.* **281,** 11515–11522 (2006).
38. Veeriah, S. *et al.* Somatic mutations of the Parkinson's disease-associated gene *PARK2* in glioblastoma and other human malignancies. *Nature Genet.* **42,** 77–82 (2010).
39. Olsen, J. H. *et al.* Atypical cancer pattern in patients with Parkinson's disease. *Br. J. Cancer* **92,** 201–205 (2005).
40. Moller, H., Mellemkjaer, L., McLaughlin, J. K. & Olsen, J. H. Occurrence of different cancers in patients with Parkinson's disease. *Br. Med. J.* **310,** 1500–1501 (1995).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements M.D.T. is the recipient of a CIHR Clinician-Scientist Phase I award, and was formerly a Sontag Distinguished Scholar with funds from the Sontag Foundation. Funding is acknowledged from the Pediatric Brain Tumour Foundation (M.D.T. and J.T.R.), and the National Institutes of Health (CA159859 to M.D.T., R.W.R. and B.W.), The Canadian Cancer Society, Genome Canada, Genome BC, Terry Fox Research Institute, Ontario Institute for Cancer Research, Pediatric Oncology Group Ontario, Funds from 'The Family of Kathleen Lorette' and the Clark H. Smith Brain Tumour Centre, Montreal Children's Hospital Foundation, Hospital for Sick Children: Sonia and Arthur Labatt Brain Tumour Research Centre, Chief of Research Fund, Cancer Genetics Program, Garron Family Cancer Centre, B.R.A.I.N. Child, CIHR (grant no. ATE-110814); the University of Toronto McLaughlin Centre, CIHR Institute of Cancer Research (grant no. AT1 – 112286) and C17 through the Advancing Technology Innovation through Discovery competition (Project Title: The Canadian Pediatric Cancer Genome Consortium: Translating next-generation sequencing technologies into improved therapies for high-risk childhood cancers). Canada's Michael Smith Genome Sciences Centre is supported by the BC Cancer Foundation. J.R. is supported by The Children's Discovery Institute. P.A.N. was supported by a Restracomp Fellowship (Hospital for Sick Children) and is currently a Roman-Herzog Postdoctoral Fellow (Hertie Foundation). Salary support for L.G. was provided by the Ontario Institute for Cancer Research through funding provided by the Government of Ontario. E.V.G.M. is supported by NIH grants CA86335, CA116804, CA138292, NCI contracts 28X5100 and 29XS193, the Southeastern Brain Tumour Foundation, and the Brain Tumour Foundation for Children. This study includes samples provided by the UK Children's Cancer and Leukaemia Group (CCLG) as part of CCLG-approved biological study BS-2007-04. J.K. and S.P. were supported by a grant from the German

Cancer Aid (109252). We thank C. Lu, K. Otaka and The Centre for Applied Genomics for technical assistance. We thank N. S. Devi and Z. Zhang for technical assistance. We thank D. Stoll for project management, S. Archer for technical writing, and P. Parouthis for artwork. The MAGIC project is part of the International Cancer Genome Consortium.

Author Contributions P.A.N. and M.D.T. co-conceived the study. P.A.N., M.A.M., and M.D.T. led the study. P.A.N. planned and executed experiments and analyses, supervised data acquisition, performed bioinformatic analyses, and extracted nucleic acids for the MAGIC cohort. D.J.H.S. led the bioinformatics and performed analyses. J.P. performed quantitative RT–PCR and Sanger sequencing of *PVT1* fusions, expression profiled *PVT1*-encoded miRNAs, and generated schematics for *PVT1* fusion genes. L.G. performed the *MYC* and *miR-1204* knockdown experiments. A.S.M. supervised the RNA-Seq and WGS experiments and performed data analysis. T.Z., A.M.S., and J.D.S. performed the large insert paired-end sequencing and PCR verification of *SNCAIP* duplication samples. A.K.o. performed interphase FISH and immunohistochemistry for candidate genes. J.R. and G.D.B. led the pathway analyses and generated enrichment plots. S.E.S. and R.B. provided technical support with the GISTIC2 bioinformatic platform. D.W.E. performed interphase FISH for candidate genes. C.R.M., A.C.L. and S.W.S. performed the SNP6 genotyping analysis, provided a database of normal copy number variants, and the control dataset used to infer copy number in the tumour samples. S.M., A.D., F.M.G.C., M.K., D.T.W.J. and H.W. performed bioinformatic analyses and provided technical advice. Y.Y. sequenced CTNNB1 in the WNT tumours. V.R., D.K., M.F.R., T.A., and P.D. performed functional assays for candidate genes. B.Lu. extracted nucleic acids, managed biobanking, and maintained the patient database. S.M. and A.R. performed the drug database analysis. Xin W., Xiaochong W. and M.R. provided technical support. R.T.C., A.C., E.D., R.G., L.H., K.L., A.L., K.L.M., K.M., A.M., L.Q.Q., A.G.J.R., N.T., R.J.V., I.B., R.A.M., A.J.M., R.H. and S.J.M.J. led the RNA-Seq and WGS experiments and performed data analyses. A.F.-L and A.M.K. provided the database of SHH-responsive genes. R.J.W.-R. W.A.G., M.P.-P., C.C.H., O.D., S.S.R., F.F.D., S.S.P.-F., B.-K.C., S.-K.K., M.-C.W., W.S., C.G.E., M.A., J.F.P., K.F., K.M.M., G.Y.G., C.C.R., L.M.M., L.M., E.M.C.M., N.K.K., P.J.F., J.M.K., J.M.O., R.G.E., K.C.T., M.K.C., B.La., R.E.M., D.D.B., A.F., S.A., N.J., J.C.L., S.B., N.G., W.A.W., L.B., A.K.I., T.E.V.M., T.K., T.T., S.K.E., J.R.L., J.B.R., J.M.L., E.G.V.M., M.F., H.N., G.C., M.G., P.H., A.G.S., A.I., S.J., C.G.C., R.V., Y.S.R., S.R., M.Z., C.C.F., J.A.C., M.L.L., Y.-J.C., U.T., C.E.H., E.B., S.W.S., J.T.R., D.M., S.C.C., S.J.M.J., J.O.K., S.M.P. and M.A.M. provided valuable input regarding study design, data analysis, and interpretation of results. P.A.N., D.J.H.S., J.P., L.G., A.S.M., A.M.K. and M.D.T. wrote the manuscript. M.A.M. and M.D.T. provided financial and technical support and oversaw the study. M.A.M. and M.D.T. are joint senior authors and project co-leaders.

Author Information SNP6 copy number and gene expression array data have been deposited at the Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/) as a GEO SuperSeries under accession number GSE37385. Whole genome and transcriptome sequencing data have been deposited at the European Genome-phenome Archive (EGA; https://www.ebi.ac.uk/ega/) hosted by the EBI, under accession number EGAD00001000158. Reprints and permissions information is available at www.nature.com/reprints. This paper is distributed under the terms of the Creative Commons Attribution-Non-Commercial-Share Alike licence, and is freely available to all readers at www.nature.com/nature. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to M.A.M. (mmarra@bcgsc.ca) or M.D.T. (mdtaylor@sickkids.ca).

Paul A. Northcott[1,2•], David J. H. Shih[1,3•], John Peacock[1,3], Livia Garzia[1], A. Sorana Morrissy[1], Thomas Zichner[4], Adrian M. Stütz[4], Andrey Korshunov[5], Jüri Reimand[6], Steven E. Schumacher[7], Rameen Beroukhim[8,9,10,11,12], David W. Ellison[13], Christian R. Marshall[14], Annaick Lionel[15], Stephen Mack[1,3], Adrian Dubuc[1,2], Yuan Yao[1,3], Vijay Ramaswamy[1,2], Betty Luu[1], Adi Rolider[1], Florence M. G. Cavalli[1], Xin Wang[1,3], Marc Remke[1,2], Xiaochong Wu[1], Readman Y. B. Chiu[16], Andy Chu[16], Eric Chuah[16], Richard D. Corbett[16], Gemma R. Hoad[16], Shaun D. Jackman[16], Yisu Li[16], Allan Lo[16], Karen L. Mungall[16], Ka Ming Nip[16], Jenny Q. Qian[16], Anthony G. J. Raymond[16], Nina Thiessen[16], Richard J. Varhol[16], Inanc Birol[16], Richard A. Moore[16], Andrew J. Mungall[16], Robert Holt[17], Daisuke Kawauchi[18], Martine F. Roussel[18], Marcel Kool[2], David T. W. Jones[2], Hendrik Witt[19,20], Andrea Fleming[21], Daniel C. Bartholomaeus[21], Kenneth C. Kennedy[21], Wechsler-Reya[22], Peter Dirks[23], Tzvi Aviv[24], Wieslawa A. Grajkowska[27], Marta Perek-Polnik[28], Christine C. Haberler[29], Olivier Delattre[30], Stéphanie S. Reynaud[31], François F. Doz[32], Sarah S. Pernet-Fattet[33], Byung-Kyu Cho[34], Seung-Ki Kim[34], Kyu-Chang Wang[34], Wolfram Scheurlen[35], Charles G. Eberhart[36], Michelle Fèvre-Montange[37], Anne Jouvet[38], Ian F. Pollack[39], Xing Fan[40], Karin M. Muraszko[41], G. Yancey Gillespie[42], Concezio Di Rocco[43], Luca Massimi[43], Erna M. C. Michiels[44], Nanne K. Kloosterhof[44,45], Pim J. French[45], Johan M. Kros[46], James M. Olson[47,48], Richard G. Ellenbogen[49], Karel Zitterbart[50,51], Leos Kren[52], Reid C. Thompson[22], Michael K. Cooper[23], Boleslaw Lach[54,55], Roger E. McLendon[56], Darell D. Bigner[56], Adam Fontebasso[57], Steffen Albrecht[58,59], Nada Jabado[57,60], Janet C. Lindsey[61], Simon Bailey[61], Nalin Gupta[62], William A. Weiss[63], László Bognár[64], Almos Klekner[64], Timothy E. Van Meter[65], Toshihiro Kumabe[66], Teiji Tominaga[66], Samer K. Elbabaa[67], Jeffrey R. Leonard[68], Joshua B. Rubin[69], Linda M. Liau[70], Erwin G. Van Meir[71], Maryam Fouladi[72], Hideo Nakamura[73], Giuseppe Cinalli[74], Miklós Garami[75], Peter Hauser[75], Ali G. Saad[76], Achille Iolascon[77,78], Shin Jung[79], Carlos G. Carlotti[80], Rajeev Vibhakar[81], Young Shin Ra[82], Shenandoah Robinson[83•], Massimo Zollo[77,78], Claudia C. Faria[84,85], Jennifer A. Chan[86], Michael L. Levy[86], Poul H. B. Sorensen[87], Matthew Meyerson[8], Scott L. Pomeroy[90], Yoon-Jae Cho[91], Gary D. Bader[6,14,92,93], Uri Tabori[94], Cynthia E. Hawkins[95], Eric Bouffet[94], Stephen W. Scherer[14,15], James T. Rutka[29], David Malkin[94,96], Steven C. Clifford[61], Steven J. M. Jones[16], Jan O. Korbel[4], Stefan M. Pfister[2,19], Marco A. Marra[17,97] & Michael D. Taylor[1,3,25]

[1]Developmental & Stem Cell Biology Program, The Hospital for Sick Children, 101 College Street, TMDT-11-401M, Toronto, Ontario M5G 1L7, Canada. [2]Division of Pediatric Neurooncology, German Cancer Research Center (DKFZ), Im Neuenheimer Feld 280, 69120 Heidelberg, Germany. [3]Department of Laboratory Medicine and Pathobiology, University of Toronto, Medical Sciences Buildings, 1 King's College Circle, 6th Floor, Toronto, Ontario M5S 1A8, Canada. [4]Genome Biology, European Molecular Biology Laboratory, Meyerhofstrasse 1, 69117 Heidelberg, Germany. [5]CCU Neuropathology, German Cancer Research Center (DKFZ), Im Neuenheimer Feld 220-221, Department of Neuropathology, University of Heidelberg, Im Neuenheimer Feld 224, 69120 Heidelberg, Germany. [6]The Donnelly Centre, University of Toronto, 160 College Street, Room 602, Toronto, Ontario M5S 3E1, Canada. [7]Department of Cancer Biology, Dana-Farber Cancer Institute, 450 Brookline Avenue, Boston, Massachusetts 02215, USA. [8]Department of Medical Oncology, Dana-Farber Cancer Institute, 450 Brookline Avenue, Boston, Massachusetts 02215, USA. [9]Department of Medicine, Harvard Medical School, 25 Shattuck Street, Boston, Massachusetts 02115, USA. [10]Department of Medicine, Brigham and Women's Hospital, 75 Francis Street, Boston, Massachusetts 02115, USA. [11]Cancer Program, Broad Institute, 7 Cambridge Center, Cambridge, Massachusetts 02142, USA. [12]Center for Cancer Genome Discovery, Dana-Farber Cancer Institute, 450 Brookline Avenue, Boston, Massachusetts 02215, USA. [13]Pathology, St Jude Children's Research Hospital, 262 Danny Thomas Place, Memphis, Tennessee 38105, USA. [14]McLaughlin Centre and Department of Molecular Genetics, University of Toronto, 101 College Street, Toronto, Ontario M5G 1L7, Canada. [15]The Centre for Applied Genomics and Program in Genetics and Genome Biology, The Hospital for Sick Children, 101 College Street, TMDT-14-701, Toronto, Ontario M5G 1L7, Canada. [16]Michael Smith Genome Sciences Centre, BC Cancer Agency, 100-570 West 7th Avenue, Vancouver, British Columbia V5Z 4S6, Canada. [17]Michael Smith Genome Sciences Centre, BC Cancer Agency, 675 West 10th Avenue, Vancouver, British Columbia V5Z 1L3, Canada. [18]Tumour Cell Biology, St Jude Children's Research Hospital, 262 Danny Thomas Place, Memphis, Tennessee 38105, USA. [19]Department of Pediatric Oncology, University Hospital Heidelberg, Im Neuenheimer Feld 430, 69120 Heidelberg, Germany. [20]Departments of Pathology and Immunology, University Hospital Heidelberg, Im Neuenheimer Feld 430, 69120 Heidelberg, Germany. [21]Pediatric Clinical Trials Office, Memorial Sloan-Kettering Cancer Center, 405 Lexington Avenue, New York, New York 10174, USA. [22]Neurological Surgery, Vanderbilt Medical Center, T-4224 MCN, Nashville, Tennessee 37232, USA. [23]Cancer Biology, Vanderbilt Medical Center, 465 21st Avenue South, MRB III 6160, Nashville, Tennessee 37232-8550, USA. [24]Sanford-Burnham Medical Research Institute, La Jolla, California 92037, USA. [25]Department of Surgery, Division of Neurosurgery and Labatt Brain Tumour Research Centre, The Hospital for Sick Children, 555 University Avenue, Hill 1503, Toronto, Ontario M5G 1X8, Canada. [26]Developmental & Stem Cell Biology Program, The Hospital for Sick Children, 101 College Street, TMDT-13-601, Toronto, Ontario M5G 1L7, Canada. [27]Department of Pathology, The Children's Memorial Health Institute, Aleja Dzieci Polskich 20, 04-730 Warsaw, Poland. [28]Department of Oncology, The Children's Memorial Health Institute, Aleja Dzieci Polskich 20, 04-730 Warsaw, Poland. [29]Institute of Neurology, Medical University of Vienna, AKH 4J, Waehringer Gürtel 18-20, A-1097 Vienna, Austria. [30]INSERM U.830, Institut Curie, 26 rue d'Ulm, 75238 Paris Cedex 5, France. [31]Unit of Somatic Genetics, Institut Curie, 26 rue d'Ulm, 75238 Paris Cedex 5, France. [32]Department of Pediatric Oncology, Institut Curie, 26 rue d'Ulm, 75248 Paris Cedex 5, France. [33]Pediatric Hematology and Oncology, CHUV University Hospital, 1011 Lausanne, Switzerland. [34]Department of Neurosurgery, Division of Pediatric Neurosurgery, Seoul National University Children's Hospital, 101 Daehak-Ro, Jongno-Gu, Seoul 110-744, South Korea. [35]Head of Pediatric Oncology, Städtische Kinderklinik, Theodor-Kutzer-Ufer 1-3, 90419 Nuremberg, Germany. [36]Departments of Pathology, Opthalmology and Oncology, John Hopkins University School of Medicine, 720 Rutland Avenue, Ross Building 558, Baltimore, Maryland 21205, USA. [37]INSERM U1028, CNRS UMR5292, Centre de Recherche en Neurosciences, Université de Lyon, 69336 Lyon, France. [38]Centre de Pathologie EST, Groupement Hospitalier EST, Université de Lyon, 69500 Bron, France. [39]Department of Neurological Surgery, University of pittsburgh School of Medicine, 4401 Penn Avenue, Pittsburgh, Pennsylvania 15224, USA. [40]Departments of Neurosurgery and Cell and Developmental Biology, University of Michigan Medical School, 109 Zina Pitcher Place, 5018 BSRB, Ann Arbor, Michigan 48109, USA. [41]Department of Neurosurgery, University of Michigan Medical School, 1500 E. Medical Center Drive, Taubman Center, Room 3552, Ann Arbor, Michigan 48109, USA. [42]Department of Surgery, Division of Neurosurgery, University of Alabama at Birmingham, 1900 University Boulevard, THT 1052, Birmingham, Alabama 35294-0006, USA. [43]Pediatric Neurosurgery, Catholic University Medical School, 00186 Rome, Italy. [44]Department of Pediatric Oncology and Hematology, Erasmus Medical Center, Dr. Molewaterplein 50, 3000 Rotterdam, The Netherlands. [45]Department of Neurology, Erasmus Medical Center, Dr. Molewaterplein 50, PO Box 2040, 3000 CA Rotterdam, The Netherlands. [46]Department of Pathology, Erasmus Medical Center, Dr. Molewaterplein 50, 3015 GE Rotterdam, The Netherlands. [47]Clinical Research Division, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, D4-100, Seattle, Washington 98109, USA. [48]Seattle Children's Hospital, Seattle, Washington 98104, USA. [49]Neurological Surgery, University of Washington School of Medicine, Harborview Medical Center, 325 Ninth Avenue, Seattle, Washington 98104, USA. [50]Department of Pediatric Oncology, School of Medicine, Masaryk University, Cernopolni 9, 613 00 Brno, Czech Republic. [51]Department of Pediatric Oncology, University Hospital Brno, 625 00 Brno, Czech Republic. [52]Department of Pathology, University Hospital Brno, Jihlavska 20, 625 00 Brno, Czech Republic. [53]Department of Neurology, Vanderbilt Medical Center, 465 21st Avenue South, MRB III 6160, Nashville, Tennessee 37232-8550, USA. [54]Department of Pathology and Molecular Medicine, Division of Anatomical Pathology, McMaster University, Hamilton, Ontario L8S 4L8, Canada. [55]Department of Pathology and Laboratory Medicine, Hamilton General Hospital, 237 Barton Street East, Hamilton, Ontario L8L 2X2, Canada. [56]Department of Pathology, Duke University, DUMC 3712,

Durham, North Carolina 27710, USA. [57]Division of Experimental Medicine, McGill University, 4060 Ste Catherine West, Montreal, Quebec H3Z 2Z3, Canada. [58]Department of Pathology, McGill University, Montreal, Quebec H3A 2B4, Canada. [59]Department of Pathology, Montreal Children's Hospital, 2300 Tupper, Montreal, Quebec H3H 1P3, Canada. [60]Department of Pediatrics, Division of Hemato-Oncology, McGill University, Montreal, Quebec H3H 1P3, Canada. [61]Northern Institute for Cancer Research, Newcastle University, Newcastle upon Tyne NE1 4LP, United Kingdom. [62]Departments of Neurological Surgery and Pediatrics, University of California San Francisco, 505 Parnassus Avenue, Room M779, San Francisco, California 94143-0112, USA. [63]Departments of Neurology, Pediatrics, and Neurosurgery, University of California San Francisco, The Helen Diller Family Cancer Research Building 1450 3rd Street, Room HD-220, MC 0520, San Francisco, California 94158, USA. [64]Department of Neurosurgery, University of Debrecen, Medical and Health Science Centre, Móricz Zs. Krt. 22., 4032 Debrecen, Hungary. [65]Pediatrics, Virginia Commonwealthy University, School of Medicine, Box 980646, Pediatric Hematology-Oncology, 1101 East Marshall Street, Richmond, Virginia 23298-0646, USA. [66]Department of Neurosurgery, Tohoku University Graduate School of Medicine, 1-1 Seiryo-machi, Aoba-ku, Sendai 980-8574, Japan. [67]Department of Neurosurgery, Division of Pediatric Neurosurgery, St Louis University School of Medicine, 1465 South Grand Boulevard, Suite 3707, St Louis, Missouri 63104, USA. [68]Department of Neurosurgery, Division of Pediatric Neurosurgery, Washington University School of Medicine and St Louis Children's Hospital, Campus Box 8057, 660 South Euclid Avenue, St Louis, Missouri 63110, USA. [69]Departments of Pediatrics, Anatomy and Neurobiology, Washington University School of Medicine and St Louis Children's Hospital, Campus Box 8208, 660 South Euclid Avenue, St Louis, Missouri 63110, USA. [70]Department of Neurosurgery, David Geffen School of Medicine at UCLA, 10833 Le Conte Avenue, Campus 690118, Los Angeles, California 90095, USA. [71]Laboratory of Molecular Neuro-Oncology, Departments of Neurosurgery and Hematology & Medical Oncology, School of Medicine and Winship Cancer Institute, Emory University, 1365C Clifton Road NE, Atlanta, Georgia 30322, USA. [72]Division of Oncology, University of Cincinnati, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio 45229, USA. [73]Department of Neurosurgery, Kumamoto University Graduate School of Medical Science, 1-1-1, Honjo, Kumamoto 860-8556, Japan. [74]Paediatric Neurosurgery, Ospedale Santobono-Pausilipon, 80145 Naples, Italy. [75]2nd Department of Pediatrics, Semmelweis University, 1085 Budapest, Hungary. [76]Pathology, University of Arkansas for Medical Sciences, 1 Children's Way, lot 820, Little Rock, Arkansas 72202, USA. [77]Dipartimento di Biochimica e Biotecnologie Mediche, University of Naples, Via Pansini 5, 80145 Naples, Italy. [78]CEINGE Biotecnologie Avanzate, Via Gaetano Salvatore 486, 80145 Naples, Italy. [79]Department of Neurosurgery, Chonnam National University Research Institute of Medical Sciences, Chonnam National University Hwasun Hospital and Medical School, 322 Seoyang-ro, Hwasun-eup, Hwasun-gun, Chonnam 519-763, South Korea. [80]Department of Surgery and Anatomy, Faculty of Medicine of Ribeirão Preto, Universidade de São Paulo, Brazil, Avenida Bandeirantes, 3900, Monte Alegre, 14049-900, Rebeirao Preto, São Paulo, Brazil. [81]Pediatrics, University of Colorado Denver, 12800 19th Avenue, Aurora, Colorado 80045, USA. [82]Department of Neurosurgery, University of Ulsan, Asan Medical Center, Seoul, 138-736, South Korea. [83]Division of Pediatric Neurosurgery, Case Western Reserve, Cleveland, Ohio 44106, USA. [84]Rainbow Babies & Children's, Cleveland, Ohio 44106, USA. [85]Division of Neurosurgery, Hospital de Santa Maria, Centro Hospitalar Lisboa Norte EPE, 1169-050, Lisbon, Portugal. [86]Cell Biology Program, The Hospital for Sick Children, 101 College Street, TMDT-401-J, Toronto, Ontario M5G 1L7, Canada. [87]Department of Pathology and Laboratory Medicine, University of Calgary, 3330 Hospital Drive NW, HRIC 2A25A, Calgary, Alberta T2N 4N1, Canada. [88]ICSD Division of Neurosurgery, Rady Children's Hospital San Diego, 8010 Frost Street, Suite 502, San Diego, California 92123, USA. [89]Department of Molecular Oncology, British Columbia Cancer Research Centre, 675 West 10th Avenue, Vancouver, British Columbia V5Z 1L3, Canada. [90]Department of Neurology, Harvard Medical School, Children's Hospital Boston, Fegan 11, 300 Longwood Avenue, Boston, Massachusetts 02115, USA. [91]Department of Neurology and Neurological Sciences, Stanford University School of Medicine, 1201 Welch Road, MSLS Building, Rm P213, Stanford, California 94305, USA. [92]Banting and Best Department of Medical Research, University of Toronto, Toronto, Ontario M5G 1L6, Canada. [93]Samuel Lunenfeld Research Institute at Mount Sinai Hospital, University of Toronto, Toronto M5G 1X5, Ontario, Canada. [94]Department of Haematology & Oncology, The Hospital for Sick Children, 555 University Avenue, Toronto, Ontario M5G 1X8, Canada. [95]Department of Pathology, The Hospital for Sick Children, 555 University Avenue, Toronto, Ontario M5G 1X8, Canada. [96]Department of Pediatrics, University of Toronto, Toronto, Ontario M5G 1X8, Canada. [97]Department of Medical Genetics, University of British Columbia, 675 West 10th Avenue, Vancouver, British Columbia V5Z 1L3, Canada.
*These authors contributed equally to this work.

LETTER

doi:10.1038/nature11361

Quantum nonlinear optics with single photons enabled by strongly interacting atoms

Thibault Peyronel[1], Ofer Firstenberg[1,2], Qi-Yu Liang[1], Sebastian Hofferberth[1,2], Alexey V. Gorshkov[3], Thomas Pohl[4], Mikhail D. Lukin[2] & Vladan Vuletić[1]

The realization of strong nonlinear interactions between individual light quanta (photons) is a long-standing goal in optical science and engineering[1,2], being of both fundamental and technological significance. In conventional optical materials, the nonlinearity at light powers corresponding to single photons is negligibly weak. Here we demonstrate a medium that is nonlinear at the level of individual quanta, exhibiting strong absorption of photon pairs while remaining transparent to single photons. The quantum nonlinearity is obtained by coherently coupling slowly propagating photons[3–5] to strongly interacting atomic Rydberg states[6–12] in a cold, dense atomic gas[13,14]. Our approach paves the way for quantum-by-quantum control of light fields, including single-photon switching[15], all-optical deterministic quantum logic[16] and the realization of strongly correlated many-body states of light[17].

Recently, remarkable advances have been made towards optical systems that are nonlinear at the level of individual photons. The most promising approaches have used high-finesse optical cavities to enhance the atom–photon interaction probability[2,18–21]. In contrast, our present method is cavity-free and is based on mapping photons onto atomic states with strong interactions in an extended atomic ensemble[13,15,22,23]. The central idea is illustrated in Fig. 1, where a quantum probe field incident onto a cold atomic gas is coupled to high-lying atomic states (Rydberg levels[24]) by means of a second, stronger laser field (control field). For a single incident probe photon, the control field induces a spectral transparency window in the otherwise opaque medium via electromagnetically induced transparency (EIT[3]), and the probe photon travels at much reduced speed in the form of a coupled excitation of light and matter (a Rydberg polariton). However, in stark contrast to conventional EIT, if two probe photons are incident onto the Rydberg medium, the strong interaction between two Rydberg atoms tunes the transition out of resonance, thereby destroying the transparency and leading to absorption[15,22,23,25,26]. The experimental demonstration of an optical material exhibiting strong two-photon attenuation in combination with single-photon transmission is the central result of this work.

The quantum nonlinearity can be viewed as a photon–photon blockade mechanism that prevents the transmission of any multi-photon state. It arises from the Rydberg excitation blockade[27], which precludes the simultaneous excitation of two Rydberg atoms that are separated by less than a blockade radius, r_b (Fig. 1). During the optical excitation under EIT conditions, an incident single photon is converted into a Rydberg polariton inside the medium. However, owing to the Rydberg blockade, a second polariton cannot travel within a blockade radius from the first one, and EIT is destroyed. Accordingly, if the second photon approaches the single Rydberg polariton, it will be significantly attenuated, provided that r_b exceeds the resonant attenuation length of the medium in the absence of EIT, $l_a = (\mathcal{N}\sigma_a)^{-1}$, where $\mathcal{N}$ is the peak atomic density and σ_a the absorption cross-section. This simple physical picture implies that, in the regime where the blockade radius exceeds the absorption length ($r_b \gtrsim l_a$), two photons in a tightly focused beam not only cannot pass through each other[15], but also cannot propagate close to each other inside the medium (see Fig. 1c and the detailed theoretical analysis below). Using Rydberg states with principal quantum numbers $46 \leq n \leq 100$, we can realize blockade radii r_b between 3 μm and 13 μm, while for our highest atomic densities of $\mathcal{N} = 2 \times 10^{12}$ cm^{-3}, the attenuation length l_a is below 2 μm. The optical medium then acts as a quantum nonlinear absorption filter, converting incident laser light into a non-classical train of single-photon pulses.

EIT nonlinearities at the few-photon level have been previously observed without using strongly interacting atomic states by means of strong transverse confinement of the light[28,29]. The interactions between cold Rydberg atoms have been explored in ensembles[6–10], and have been used to realize quantum logic gates between two Rydberg atoms[11,12,24]. Enhanced optical nonlinearities using Rydberg EIT[15,22,23] have been observed in pioneering work[13,14] that we are building on. Very recently, the Rydberg blockade in a dense, mesoscopic atomic ensemble has been used to implement a deterministic single-photon source[30].

To observe the photon–photon blockade, several key requirements must be fulfilled. First, to eliminate Doppler broadening, the atoms should be cold so that they move by less than an optical wavelength during the microsecond lifetime of the EIT coherence. Second, the atomic cloud should be sufficiently dense such that the blockade condition $r_b \gtrsim l_a$ is fulfilled. Last, the system should be one-dimensional, that is, the transverse size of the probe beam should be smaller than the blockade radius r_b in order to prevent polaritons from travelling side by side.

We fulfil these conditions by trapping a dense laser-cooled atomic ensemble and focusing the probe beam to a Gaussian waist $w = 4.5$ μm $< r_b$ (see Methods). We prepare an ^{87}Rb ensemble containing up to 10^5 atoms in a far-detuned optical dipole trap produced by a Nd:YAG laser operating at 1,064 nm with a total power of 5 W. The trap is formed by two orthogonally polarized beams with waists $w_t = 50$ μm intersecting at an angle of 32°. The atoms are optically pumped into the hyperfine (F) and magnetic (m_F) sublevel $|g\rangle = |5S_{1/2}, F = 2, m_F = 2\rangle$ in the presence of a 3.6 G magnetic field along the quantization axis, which is defined by the common propagation direction of the probe and control beams along the long axis of the cloud. The probe beam on the $|g\rangle \rightarrow |e\rangle = |5P_{3/2}, F = 3, m_F = 3\rangle$ transition and the control beam on the $|e\rangle \rightarrow |r\rangle = |nS_{1/2}, J = 1/2, m_J = 1/2\rangle$ transition with waist $w_c = 12.5$ μm are oppositely circularly polarized. (Here J and m_J denote the quantum numbers for total angular momentum and its component along the quantization axis, respectively.) The resonant optical depth (OD) of the cloud can be as large as OD = 50, with in-trap radial and axial r.m.s. cloud dimensions of $\sigma_r = 10$ μm and $\sigma_{ax} = 36$ μm, respectively. To avoid inhomogeneous light-shift broadening of the two-photon transition, we turn off the

[1]Department of Physics and Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. [2]Department of Physics, Harvard University, Cambridge, Massachusetts 02138, USA. [3]Institute for Quantum Information and Matter, California Institute of Technology, Pasadena, California 91125, USA. [4]Max Planck Institute for the Physics of Complex Systems, Nöthnitzer Strasse 38, D-01187 Dresden, Germany.

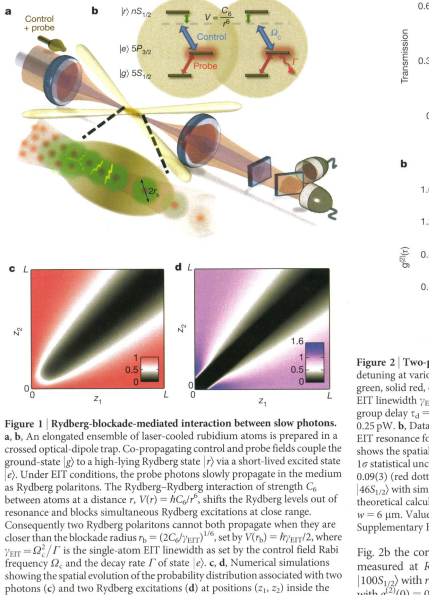


Figure 1 | Rydberg-blockade-mediated interaction between slow photons. **a**, **b**, An elongated ensemble of laser-cooled rubidium atoms is prepared in a crossed optical-dipole trap. Co-propagating control and probe fields couple the ground-state $|g\rangle$ to a high-lying Rydberg state $|r\rangle$ via a short-lived excited state $|e\rangle$. Under EIT conditions, the probe photons slowly propagate in the medium as Rydberg polaritons. The Rydberg–Rydberg interaction of strength C_6 between atoms at a distance r, $V(r) = \hbar C_6/r^6$, shifts the Rydberg levels out of resonance and blocks simultaneous Rydberg excitations at close range. Consequently two Rydberg polaritons cannot both propagate when they are closer than the blockade radius $r_b = (2C_6/\gamma_{EIT})^{1/6}$, set by $V(r_b) = \hbar\gamma_{EIT}/2$, where $\gamma_{EIT} = \Omega_c^2/\Gamma$ is the single-atom EIT linewidth as set by the control field Rabi frequency Ω_c and the decay rate Γ of state $|e\rangle$. **c**, **d**, Numerical simulations showing the spatial evolution of the probability distribution associated with two photons (**c**) and two Rydberg excitations (**d**) at positions (z_1, z_2) inside the medium, normalized by their values in the absence of blockade. Two Rydberg excitations are excluded from the blockaded range, resulting in the formation of an anti-bunching feature in the light field, whose width increases during the propagation due to the finite EIT transparency width $B = \gamma_{EIT}/\sqrt{8OD}$.

optical dipole trap before turning on the probe and control light, and probe the Rydberg EIT system continuously for up to a few hundred microseconds. The control light is filtered out from the transmitted light, and the photon–photon correlation function $g^{(2)}(\tau)$ of the probe beam is measured versus the time separation τ by means of two photon counters. The slow-light group delay τ_d through the atomic medium[5] is measured independently in a pulsed experiment, and used to calculate the corresponding minimum group velocity, $v_g = \sqrt{2}\pi\sigma_{ax}/\tau_d$.

Probe transmission spectra are presented in Fig. 2a for large optical depth OD = 40 and the control laser tuned to the Rydberg state $|100S_{1/2}\rangle$. At very low incident photon rates $R_i \leq 1\ \mu s^{-1}$, the spectrum displays an EIT transparency window with 60% transmission. The transmission is mainly limited by the finite EIT decoherence rate γ_{gr}, which for our system is dominated by Doppler broadening and laser linewidth. The extraordinary nonlinearity of the Rydberg EIT medium[13] becomes apparent as the incident photon rate is increased: the probe beam is already strongly attenuated at a photon rate of $R_i \approx 4\ \mu s^{-1}$. To demonstrate that we are operating in a quantum nonlinear regime, we show in

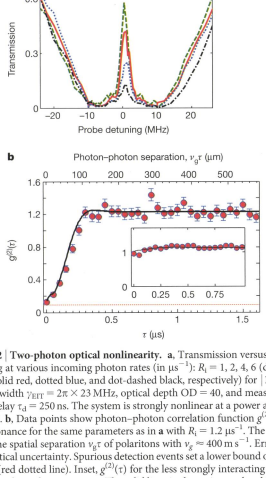


Figure 2 | Two-photon optical nonlinearity. **a**, Transmission versus probe detuning at various incoming photon rates (in μs^{-1}): $R_i = 1, 2, 4, 6$ (dashed green, solid red, dotted blue, and dot-dashed black, respectively) for $|100S_{1/2}\rangle$, EIT linewidth $\gamma_{EIT} = 2\pi \times 23$ MHz, optical depth OD = 40, and measured group delay $\tau_d = 250$ ns. The system is strongly nonlinear at a power as low as 0.25 pW. **b**, Data points show photon–photon correlation function $g^{(2)}(\tau)$ at EIT resonance for the same parameters as in **a** with $R_i = 1.2\ \mu s^{-1}$. The top axis shows the spatial separation $v_g\tau$ of polaritons with $v_g \approx 400$ m s^{-1}. Error bars, 1σ statistical uncertainty. Spurious detection events set a lower bound on $g^{(2)}$ of 0.09(3) (red dotted line). Inset, $g^{(2)}(\tau)$ for the less strongly interacting state $|46S_{1/2}\rangle$ with similar parameters. The solid lines in the main panel and inset are theoretical calculations as described in the text, with the probe waist fixed at $w = 6$ μm. Values $g^{(2)} > 1$ are attributed to classical fluctuations (see Supplementary Fig. 4 and Supplementary Information).

Fig. 2b the correlation function $g^{(2)}(\tau)$ of the transmitted probe light, measured at $R_i = 1.2\ \mu s^{-1}$. For the most strongly interacting state $|100S_{1/2}\rangle$ with $r_b = 13\ \mu m \approx 5l_a \approx 2.9w$ we observe strong antibunching with $g^{(2)}(0) = 0.13(2)$, largely limited by background light. (Here 0.13(2) indicates 0.13 ± 0.02.) Subtraction of the independently measured background coincidence counts yields a corrected $g_c^{(2)}(0) = 0.04(3)$. These observations are in sharp contrast to EIT transmission via a less strongly interacting Rydberg state $|46S_{1/2}\rangle$ with $r_b = 3\ \mu m$, where the photon statistics of the transmitted light are similar to those of the incident coherent state (see Fig. 2b inset). We note that for $|100S_{1/2}\rangle$ the photons are anti-bunched over a length scale $v_g\tau \approx 50\ \mu m$ that exceeds the blockade radius (see top axis of Fig. 2b), indicating the influence of additional propagation effects beyond the simple picture outlined above.

To investigate the transmission characteristics of multiple photons through the medium, we plot in Fig. 3a the output photon rate R_o, scaled by the EIT transmission measured at low probe power, as a function of incident photon rate R_i. At first, R_o increases linearly with R_i as expected, but then saturates abruptly to a constant value of $R_o = 1.3(3)\ \mu s^{-1}$. Note that these observations deviate from the simplistic model of a multiphoton absorber that transmits only the one-photon component from the incoming coherent state (black dashed line in Fig. 3a). At the same time, the observed output flux corresponds to less than one photon in the medium ($R_o^{-1} > \tau_d = 300$ ns). Figure 3b shows the saturated output rate versus the ratio r_b/w of blockade radius and probe beam waist for a wide range of principal quantum numbers, control field intensities and optical depths. The approximate

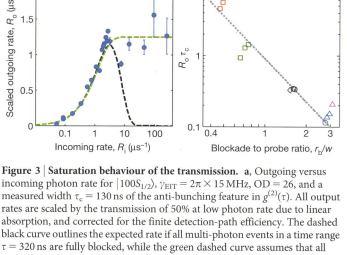


Figure 3 | Saturation behaviour of the transmission. a, Outgoing versus incoming photon rate for $|100S_{1/2}\rangle$, $\gamma_{EIT} = 2\pi \times 15$ MHz, OD = 26, and a measured width $\tau_c = 130$ ns of the anti-bunching feature in $g^{(2)}(\tau)$. All output rates are scaled by the transmission of 50% at low photon rate due to linear absorption, and corrected for the finite detection-path efficiency. The dashed black curve outlines the expected rate if all multi-photon events in a time range $\tau = 320$ ns are fully blocked, while the green dashed curve assumes that all multiphoton states within $\tau = 800$ ns are converted into an outgoing one-photon state. **b**, Saturated rate of outgoing photons $R_b\tau_c$ per anti-bunching correlation time τ_c, scaled by the linear absorption, as a function of the ratio between the blockade radius r_b and the probe beam waist w. The Rydberg states are $|100S_{1/2}\rangle$ (blue, $w = 4.5$ μm, $OD_b \approx 8$; pink, $w = 4.5$ μm, $OD_b \approx 4$), $|77S_{1/2}\rangle$ (black, $w = 4.5$ μm, $OD_b \approx 3$), $|46S_{1/2}\rangle$ (green, $w = 4.5$ μm, $OD_b \approx 0.7$; red, $w = 7$ μm, $OD_b \approx 0.7$). τ_c is estimated using the single-atom EIT linewidths (squares, triangles, circles, diamonds) $= 2\pi \times$ (6–16,18–26,29–36,50) MHz, and varies from 60 to 330 ns. The dashed line corresponds to $0.9(w/r_b)^2$, indicating the expected scaling with transverse confinement for $w \gtrsim r_b$.

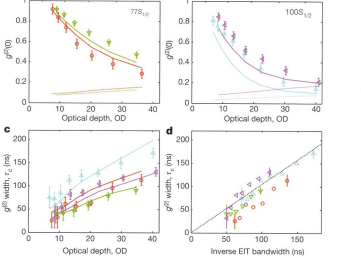


Figure 4 | Dependence of the correlation function on EIT parameters. **a**, **b**, Equal-time photon–photon correlation $g^{(2)}(0)$ as a function of OD for $|77S_{1/2}\rangle$ (**a**) and $|100S_{1/2}\rangle$ (**b**), for a set of single-atom EIT linewidths (circles, downtriangles, uptriangles, left-triangles) $= 2\pi \times$ (20,27,16,26) MHz. **c**, **d**, Width τ_c of the anti-bunching feature in $g^{(2)}(\tau)$ as a function of optical depth (**c**) and EIT transparency width (**d**) $B = \gamma_{EIT}/\sqrt{8OD}$, respectively. Solid lines in **a**–**c** are numerical solutions for a probe beam waist $w = 6$ μm, including detection noise in **a**, **b** (dotted lines). The black dashed line in **d** is $1.05/B$ and derives from an approximate analytical solution of equation (1) (see Supplementary Information). Error bars, 1σ statistical uncertainty.

$R_b \propto (w/r_b)^2$ scaling, valid for $w \gtrsim r_b$, indicates that the saturated rate for intermediate to strong interactions, $r_b \gtrsim l_a$, is largely determined by the transverse geometrical constraint, that is, by the extent to which the Rydberg polaritons can propagate side by side through the medium.

Two important features of the photon–photon blockade are the degree of two-photon suppression at equal times, $g^{(2)}(0)$, and the associated correlation time, that is, the width τ_c of the antibunching feature in $g^{(2)}(\tau)$. As discussed in detail below, the blockade mechanism is most effective if the optical depth per blockade radius, $OD_b = r_b/l_a$, exceeds unity[15], and if the system is effectively one-dimensional, $r_b > w$. Because the blockade radius[27] increases with the principal quantum number n as $r_b \propto n^{11/6}$, the combination of both effects results in a steep dependence of $g^{(2)}(0)$ upon n. Figure 4a, b shows that $g^{(2)}(0)$ improves with the principal quantum number n of the Rydberg state and the interaction strength r_b/l_a, resulting in a more than tenfold suppression of the two-photon transmission, limited by independently measured background light on the photon detectors (dotted lines). At the same time, the observed width τ_c of the $g^{(2)}$ feature considerably exceeds the photon travel time $\tau_b = r_b/v_g \approx 50$ ns through the blockade radius (Fig. 4c, d). Close examination (Fig. 4d) reveals that the correlation time is of the same order as, and scales inversely proportionally with, the spectral width[5] $B = \gamma_{EIT}/\sqrt{8OD}$ of the EIT transparency window. This observation suggests that propagation effects play an important role in establishing the $g^{(2)}$ correlation time τ_c in a medium of large optical depth. We observe that, under appropriate conditions, two-photon events are suppressed inside the medium on a length scale that approaches the size $\sigma_{ax} \approx 40$ μm of the entire atomic ensemble, and on a timescale that approaches the intrinsic coherence time $\gamma_{gr}^{-1} = 500$ ns.

To gain further understanding of these observations, we theoretically analyse the photon propagation dynamics in the weak-probe limit where the average number of photons inside the medium is much less than one. In this case, it suffices to consider two polaritons (Fig. 1b). The corresponding two-photon component of the state vector[15] is

$$|\psi_2(t)\rangle = \frac{1}{2}\int d\mathbf{r}_1 d\mathbf{r}_2 EE(\mathbf{r}_1,\mathbf{r}_2,t)\hat{\varepsilon}^\dagger(\mathbf{r}_1)\hat{\varepsilon}^\dagger(\mathbf{r}_2)|0\rangle,\ \text{where}\ \hat{\varepsilon}(\mathbf{r})$$

denotes the photon field operator, and $|EE(\mathbf{r}_1,\mathbf{r}_2,t)|^2$ is the probability of finding two photons at locations $\mathbf{r}_1$, $\mathbf{r}_2$. This probability directly yields the spatially dependent photon–photon correlation function, and, via the group velocity v_g, the corresponding temporal correlation function $g^{(2)}(\tau)$. An intuitive picture emerges if we make the simplification of a tightly focused probe beam (one-dimensional approximation) travelling through a homogeneous medium with perfect linear EIT transmission. In this case, the steady-state two-photon amplitude in the medium obeys (see Supplementary Information):

$$\partial_r EE(z_1,z_2) = -\frac{\mathcal{V}(r)}{l_a} EE(z_1,z_2) + 4l_a\left[1+\mathcal{V}(r)\frac{\Omega_c^2}{\Gamma^2}\right]\partial_r^2 EE(z_1,z_2)\,(1)$$

where $R = (z_1 + z_2)/2$ and $r = z_1 - z_2$ are the centre-of-mass and relative coordinates of the two photons, respectively. The function $\mathcal{V}(r) = r_b^6/(r_b^6 - 2ir^6)$ can be regarded as an effective potential that describes the impact of Rydberg–Rydberg interactions[15]. For large photon–photon distances, $r \gg r_b$, the potential $\mathcal{V}$ vanishes, and equation (1) yields perfect transmission under EIT, while for distances $r \lesssim r_b$, the interaction $\mathcal{V}$ modifies the two-photon propagation. According to equation (1), photon correlations emerge from a combination of two processes: the first term acts inside the blockade radius r_b, and describes absorption with a coefficient l_a^{-1} as the interaction $\mathcal{V}$ tunes EIT out of resonance. This would create a sharp dip in the two-photon correlation function with a corresponding correlation time $\tau_b = r_b/v_g$ associated with the blockade radius. However, if the corresponding spectral width $\sim \tau_b^{-1}$ exceeds the spectral width B of the EIT transparency window[5], the second diffusion-like term acts to broaden the absorption dip (Fig. 1c) in space and time, increasing the photon–photon correlation time τ_c beyond τ_b towards a value set by the EIT transparency width (Fig. 4d). To maintain strong two-photon suppression ($g^{(2)}(0) \ll 1$) in the presence of EIT-induced diffusion, the loss term must exceed the diffusion on the length scale of the blockade radius, requiring $r_b > l_a$. Large optical depth $OD_b = r_b/l_a$ of the blockaded region is therefore the

key experimental feature that allows us to extend the earlier studies[13,14] into the quantum nonlinear regime.

For direct comparisons with our experiments, we solve numerically the full set of propagation equations accounting for the Gaussian density profile of the trapped atomic cloud, the finite waist of the probe beam, and the imperfect single-photon transmission due to finite decoherence γ_{gr} of the two-photon transition. As shown in Figs 2 and 4, the theory captures the essential features of our measured correlation functions and, moreover, reproduces their dependence on the Rydberg states, control laser intensities and optical depths of the sample over a wide range of parameters. This detailed theoretical understanding also allows us to analyse the prospects for possible future improvements. These include a reduction of Doppler broadening (through lower atomic temperature or the use of counter-propagating probe and control beams) to increase the linear transmission from 60% towards unity, the excitation of even higher-lying Rydberg states for larger blockade radius, and larger atomic density to further increase the optical depth per blockade radius OD_b and overall optical depth OD.

Our observations suggest intriguing prospects for ultimate quantum control of light quanta. For example, by storing a single photon in a Rydberg state and subsequently transmitting a second Rydberg polariton, a single-photon switch can be created[15]. It can be used, for example, for quantum non-demolition measurements of optical photons. At the same time, by using strong interactions in the dispersive regime, the present approach can be used to implement deterministic quantum logic gates[15,16], which would constitute a major advance towards all-optical quantum information processing[11]. Last, our results may open the door to exploring the quantum dynamics of strongly interacting photonic many-body systems. For example, it may be possible to create a crystalline state of strongly interacting polaritons[17]. Beyond these specific applications, our work demonstrates that unique quantum nonlinear optical materials can be created by combining slow-light propagation with strong atom–atom interactions, an approach which can be potentially extended to realize other material systems with quantum nonlinearities.

METHODS SUMMARY

An ensemble of 6×10^6 laser-cooled atoms is captured in a magneto-optical trap (MOT) every 300 ms. The trapped cloud is compressed and loaded into the dipole trap by the combined actions of increasing the magnetic-field gradient to 35 G cm^{-1}, detuning the MOT trapping frequency by -30 MHz and reducing the MOT repumper intensity to 10 μW cm^{-2}. The magnetic fields are then rapidly shut off, allowing for 10 ms of molasses cooling to a temperature of 35 μK. The crossed dipole trap holds up to 10^5 atoms at a peak density of 2×10^{12} atoms cm^{-3} and a measured optical depth of OD = 50.

The probe beam is focused to a $1/e^2$ waist $w = 4.5$ μm by a confocal arrangement of achromatic doublet lenses with focal length 30 mm and diameter 6.25 mm. The control field is co-propagating with the probe beam. The frequencies of both lasers are locked to an optical Fabry–Perot resonator that is stabilized against long-term drifts to a Doppler-free atomic resonance line. The measured short-term linewidths are 120 kHz and 80 kHz for the probe and control laser, respectively. The transmitted control light is separated from the probe light by a combination of interference and absorption filters.

The intensity correlation function of the outgoing probe field is measured with two single-photon detectors. Spurious detection events typically limit $g_2(\tau)$ to ≥ 0.1. These include dark counts from the detector, imperfect polarization of the probe photons (light with the orthogonal circular polarization is only weakly absorbed by the medium) and residual control light.

Received 28 May; accepted 28 June 2012.
Published online 25 July 2012.

1. Yamamoto, Y. & Imamoglu, A. *Mesoscopic Quantum Optics* (Wiley & Sons, 1999).
2. Birnbaum, K. M. *et al.* Photon blockade in an optical cavity with one trapped atom. *Nature* **436,** 87–90 (2005).
3. Hau, L. V., Harris, S. E., Dutton, Z. & Behroozi, C. H. Light speed reduction to 17 metres per second in an ultracold atomic gas. *Nature* **397,** 594–598 (1999).
4. Fleischhauer, M. & Lukin, M. D. Dark-state polaritons in electromagnetically induced transparency. *Phys. Rev. Lett.* **84,** 5094–5097 (2000).
5. Fleischhauer, M., Imamoglu, A. & Marangos, J. P. Electromagnetically induced transparency: optics in coherent media. *Rev. Mod. Phys.* **77,** 633–673 (2005).
6. Tong, D. *et al.* Local blockade of Rydberg excitation in an ultracold gas. *Phys. Rev. Lett.* **93,** 063001 (2004).
7. Singer, K., Reetz-Lamour, M., Amthor, T., Marcassa, L. G. & Weidemüller, M. Suppression of excitation and spectral broadening induced by interactions in a cold gas of Rydberg atoms. *Phys. Rev. Lett.* **93,** 163001 (2004).
8. Liebisch, T. C., Reinhard, A., Berman, P. R. & Raithel, G. Atom counting statistics in ensembles of interacting Rydberg atoms. *Phys. Rev. Lett.* **95,** 253002 (2005).
9. Heidemann, R. *et al.* Rydberg excitation of Bose-Einstein condensates. *Phys. Rev. Lett.* **100,** 033601 (2008).
10. Johnson, T. A. *et al.* Rabi oscillations between ground and Rydberg states with dipole-dipole atomic interactions. *Phys. Rev. Lett.* **100,** 113003 (2008).
11. Urban, E. *et al.* Observation of Rydberg blockade between two atoms. *Nature Phys.* **5,** 110–114 (2009).
12. Gaëtan, A. *et al.* Observation of collective excitation of two individual atoms in the Rydberg blockade regime. *Nature Phys.* **5,** 115–118 (2009).
13. Pritchard, J. D. *et al.* Cooperative atom-light interaction in a blockaded Rydberg ensemble. *Phys. Rev. Lett.* **105,** 193603 (2010).
14. Pritchard, J. D., Weatherill, K. J. & Adams, C. S. Non-linear optics using cold Rydberg atoms. Preprint at http://arXiv.org/abs/1205.4890v1 (2012).
15. Gorshkov, A. V., Otterbach, J., Fleischhauer, M., Pohl, T. & Lukin, M. D. Photon-photon interactions via Rydberg blockade. *Phys. Rev. Lett.* **107,** 133602 (2011).
16. Shahmoon, E., Kurizki, G., Fleischhauer, M. & Petrosyan, D. Strongly interacting photons in hollow-core waveguides. *Phys. Rev. A* **83,** 033806 (2011).
17. Chang, D. E. *et al.* Crystallization of strongly interacting photons in a nonlinear optical fibre. *Nature Phys.* **4,** 884–889 (2008).
18. Schuster, I. *et al.* Nonlinear spectroscopy of photons bound to one atom. *Nature Phys.* **4,** 382–385 (2008).
19. Fushman, I. *et al.* Controlled phase shifts with a single quantum dot. *Science* **320,** 769–772 (2008).
20. Reinhard, A. *et al.* Strongly correlated photons on a chip. *Nature Photon.* **6,** 93–96 (2011).
21. Tanji-Suzuki, H., Chen, W., Landig, R., Simon, J. & Vuletić, V. Vacuum-induced transparency. *Science* **333,** 1266–1269 (2011).
22. Petrosyan, D., Otterbach, J. & Fleischhauer, M. Electromagnetically induced transparency with Rydberg atoms. *Phys. Rev. Lett.* **107,** 213601 (2011).
23. Sevinçli, S., Henkel, N., Ates, C. & Pohl, T. Nonlocal nonlinear optics in cold Rydberg gases. *Phys. Rev. Lett.* **107,** 153001 (2011).
24. Saffman, M., Walker, T. G. & Mølmer, K. Quantum information with Rydberg atoms. *Rev. Mod. Phys.* **82,** 2313–2363 (2010).
25. Møller, D., Madsen, L. B. & Mølmer, K. Quantum gates and multiparticle entanglement by Rydberg excitation blockade and adiabatic passage. *Phys. Rev. Lett.* **100,** 170504 (2008).
26. Müller, M., Lesanovsky, I., Weimer, H., Büchler, H. P. & Zoller, P. Mesoscopic Rydberg gate based on electromagnetically induced transparency. *Phys. Rev. Lett.* **102,** 170502 (2009).
27. Lukin, M. D. *et al.* Dipole blockade and quantum information processing in mesoscopic atomic ensembles. *Phys. Rev. Lett.* **87,** 037901 (2001).
28. Bajcsy, M. *et al.* Efficient all-optical switching using slow light within a hollow fiber. *Phys. Rev. Lett.* **102,** 203902 (2009).
29. Venkataraman, V., Saha, K., Londero, P. & Gaeta, A. L. Few-photon all-optical modulation in a photonic band-gap fiber. *Phys. Rev. Lett.* **107,** 193902 (2011).
30. Dudin, Y. O. & Kuzmich, A. Strongly interacting Rydberg excitations of a cold atomic gas. *Science* **336,** 887–889 (2012).
31. Han, Y., He, B., Heshami, K., Li, C.-Z. & Simon, C. Quantum repeaters based on Rydberg-blockade-coupled atomic ensembles. *Phys. Rev. A* **81,** 052311 (2010).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We acknowledge technical support from A. Mazurenko. This work was supported in part by NSF, CUA and the AFOSR Quantum Memories MURI. A.V.G. acknowledges funding from the Lee A. DuBridge Foundation and the IQIM, an NSF Physics Frontiers Center with support from the Gordon and Betty Moore Foundation.

Author Contributions The experiment was designed and built by S.H., T. Peyronel and Q.-Y.L. Measurements and analysis of the data presented were carried out by T. Peyronel, O.F. and Q.-Y.L. The theoretical analysis was performed by A.V.G. and T. Pohl. All experimental and theoretical work was supervised by M.D.L. and V.V. All authors discussed the results and contributed to the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to M.D.L. (lukin@fas.harvard.edu) and V.V. (vuletic@mit.edu).

LETTER

doi:10.1038/nature11265

Quantum phase transition in a resonant level coupled to interacting leads

Henok T. Mebrahtu[1], Ivan V. Borzenets[1], Dong E. Liu[1], Huaixiu Zheng[1], Yuriy V. Bomze[1], Alex I. Smirnov[2], Harold U. Baranger[1] & Gleb Finkelstein[1]

A Luttinger liquid is an interacting one-dimensional electronic system, quite distinct from the 'conventional' Fermi liquids formed by interacting electrons in two and three dimensions[1]. Some of the most striking properties of Luttinger liquids are revealed in the process of electron tunnelling. For example, as a function of the applied bias voltage or temperature, the tunnelling current exhibits a non-trivial power-law suppression[2,3]. (There is no such suppression in a conventional Fermi liquid.) Here, using a carbon nanotube connected to resistive leads, we create a system that emulates tunnelling in a Luttinger liquid, by controlling the interaction of the tunnelling electron with its environment. We further replace a single tunnelling barrier with a double-barrier, resonant-level structure and investigate resonant tunnelling between Luttinger liquids. At low temperatures, we observe perfect transparency of the resonant level embedded in the interacting environment, and the width of the resonance tends to zero. We argue that this behaviour results from many-body physics of interacting electrons, and signals the presence of a quantum phase transition[4,5]. Given that many parameters, including the interaction strength, can be precisely controlled in our samples, this is an attractive model system for studying quantum critical phenomena in general, with wide-reaching implications for understanding quantum phase transitions in more complex systems, such as cold atoms[6] and strongly correlated bulk materials[7].

Unlike two- and three-dimensional Fermi liquids, a Luttinger liquid completely 'dissolves' individual electrons, replacing them with collective plasmon waves. When, in the process of quantum-mechanical tunnelling, an external electron is added to the Luttinger liquid, the plasmons spread the charge through the system, akin to the ripples from a raindrop on the surface of a pond. At zero temperature, the tunnelling electron does not have the necessary energy to excite the plasmons. As a result, the tunnelling conductance between a normal metal and a Luttinger liquid, or between two Luttinger liquids, is suppressed at low temperature, with a power-law dependence on temperature[2,3].

Even more interesting is the case of resonant tunnelling, in which a single tunnel barrier between Luttinger liquids is replaced by a resonant level formed in a double-barrier quantum structure. Starting with the seminal work of ref. 8, this problem has received significant theoretical attention[9–12]. Perhaps the most spectacular prediction is the existence of resonance peaks with perfect conductance (full transparency), but vanishingly small width (infinite lifetime) at zero temperature[8]. These resonances require two Luttinger liquids that are symmetrically coupled to the resonant level. Several experiments have addressed resonant tunnelling in a Luttinger liquid in the low-temperature limit[2,13,14], but with no attempt to control the tunnelling strength. In this work, we create a system analogous to a Luttinger liquid by properly designing electron interactions in the resonant level's environment, and we tune the system to the symmetric coupling point.

We realize both the single-barrier tunnelling and the double-barrier resonant tunnelling regimes in short (~300-nm) segments of carbon nanotubes (Fig. 1a). Fig. 1b shows the representative electrical conductance through such a sample. The size quantization of electron states in the nanotube, combined with the mutual repulsion of the electrons, gives rise to a 'Coulomb blockade' pattern[15,16].

We first show Luttinger-liquid-like properties in tunnelling through a single barrier by tuning the gate voltage to Coulomb blockade valleys Y and Z of Fig. 1b, where no low-energy excitations exist in the nanotube. Electrons are then transmitted through the nanotube by co-tunnelling processes (see Supplementary Information section I). These processes are almost independent of energy, at scales smaller than the nanotube charging energy and level spacing (both millielectronvolts); at such energies, the nanotube should behave just like a single tunnel barrier.

The conductance in valleys Y and Z, plotted against bias voltage, V, shows a surprising zero-bias anomaly (ZBA), which gets progressively

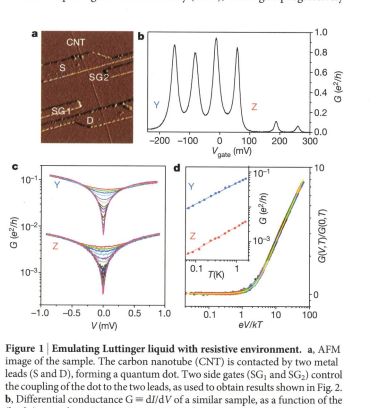


Figure 1 | Emulating Luttinger liquid with resistive environment. a, AFM image of the sample. The carbon nanotube (CNT) is contacted by two metal leads (S and D), forming a quantum dot. Two side gates (SG_1 and SG_2) control the coupling of the dot to the two leads, as used to obtain results shown in Fig. 2. **b**, Differential conductance $G \equiv dI/dV$ of a similar sample, as a function of the (back-)gate voltage, V_{gate}, at $T = 1.8$ K. We focus on Coulomb blockade valleys Y and Z, in which electron transport is conducted through co-tunnelling processes. **c**, Differential conductance versus bias voltage V, showing a pronounced zero-bias anomaly in both valleys. $T = 1.7$–0.03 K (top to bottom). **d**, $G(V,T)$ data measured in valley Y at different temperatures (coloured points) can be rescaled to collapse on the same universal curve, described by the theoretical expression of ref. 21 (yellow line). Inset shows the clear power-law dependence of zero-bias conductance $G(0,T)$ on temperature, with the same exponent in both valleys.

[1]Department of Physics, Duke University, Durham, North Carolina 27708, USA. [2]Department of Chemistry, North Carolina State University, Raleigh, North Carolina 27695, USA.

deeper as the temperature decreases (Fig. 1c). As the shape of the ZBA in the two valleys is the same up to an overall scale factor, the existence of the ZBA is not due to the nanotube itself. Indeed, the distinct feature of our samples is the metal leads to the nanotube, which are made rather resistive (kilohms; see Supplementary Information section II). 'Tunnelling with dissipation'[17] between such resistive leads is known to result in suppressed conductance: $dI/dV \equiv G \propto \max(k_B T, eV)^{2r}$ (refs 18–21), where I is current, G is conductance, k_B is the Boltzmann constant, T is temperature and $r = e^2R/h$, the ratio of the resistance of the leads, R, to the quantum resistance h/e^2. This expression is similar to the power-law suppression expected for tunnelling in a Luttinger liquid[2,3]. However, no real Luttinger liquid is present in our nanotube; at our operating temperatures, the length of an ideal clean nanotube would have to be about 100 μm to suppress the size quantization. Note the highly unusual appearance of the resistance in the exponent, which allows us to control the strength of tunnelling suppression simply by changing R.

Experimentally, the zero-bias conductance scales as $G(0,T) \propto T^{2r}$, with the same exponent $2r \approx 0.6$ found in both valleys (Fig. 1d inset); this value is consistent with the leads resistance ($R \approx 6.5\ \mathrm{k\Omega}$ in this sample; see Supplementary Information section II). Furthermore, we can rescale the whole set of $G(V,T)$ curves measured in valley Y as shown in Fig. 1d, which presents $G(V,T)/G(0,T)$ as a function of eV/k_BT —a dimensionless ratio of bias to temperature. The yellow curve overlying the symbols is the result of the full theoretical expression describing tunnelling with dissipation[21], in which we use the same value of $r = 0.3$ extracted from the temperature dependence.

The expression used to fit the data in Fig. 1d is similar to the one describing tunnelling between two Luttinger liquids[22]. The similarity may be understood qualitatively: for both tunnelling in a dissipative environment and in a Luttinger liquid, the tunnelling electron's charge couples to a continuum of bosonic modes (plasmons); at zero energy (temperature or bias), the electron cannot excite the modes, and tunnelling is suppressed. Furthermore, the formal mapping of the two problems has been demonstrated for the single-barrier case in ref. 23. The recipe is to replace the Luttinger interaction parameter g by $1/(r+1)$: *for example*, the case of vanishing dissipation, $r = 0$, corresponds to the non-interacting Luttinger liquid, $g = 1$. It is important to realize that for $r \neq 0$ electrons do in fact interact with each other through their coupling to the bosonic modes. We use the analogy between tunnelling in a dissipative environment and tunnelling in a Luttinger liquid through the rest of this text.

Having established Luttinger-liquid-like behaviour in single-barrier tunnelling, we now turn to the main focus of this work: resonant tunnelling between interacting leads. We study single-electron conductance peaks, similar to those shown in Fig. 1b, but measured on a different sample. A key feature of our experiment is the use of additional side gates to tune the coupling of the resonant level to the leads (Fig. 1a). Figure 2a shows the differential conductance map as a function of the side- and back-gate voltages. Clearly, the heights of the peaks change along the traces. For several of the peaks, the conductance reaches a maximum at some intermediate value of the side-gate voltage, indicating that the tunnelling rates from the resonant level to the source and the drain are equal: $\Gamma_S = \Gamma_D$ ('symmetric coupling').

We focus on peak X of Fig. 2a, with the side gate tuned so that the tunnelling is either symmetric (Fig. 2b) or asymmetric (Fig. 2c). Clearly, the two cases behave in markedly different ways. In the asymmetric case, the peak height decreases at low temperatures, while the width saturates[24]. In the symmetric case, the peak width decreases, while the peak height grows and reaches e^2/h. It is remarkable that the resonant tunnelling conductance can reach the unitary limit despite coupling to the interacting leads, which suppress tunnelling in the single-barrier case.

To account for the observed behaviour, we have developed a model (see Supplementary Information sections IV–VI) of a resonant level connected to two electron reservoirs, with excitation of environmental

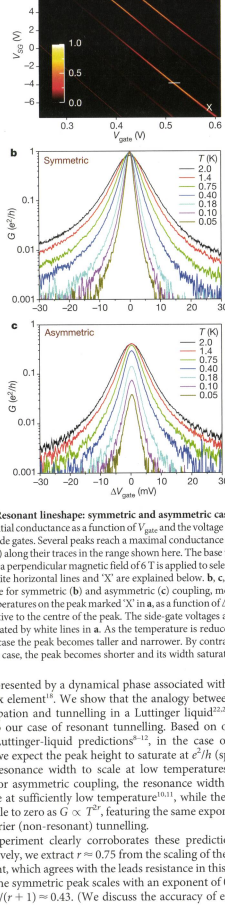


Figure 2 | Resonant lineshape: symmetric and asymmetric cases. a, Zero-bias differential conductance as a function of V_{gate} and the voltage V_{SG} applied to one of the side gates. Several peaks reach a maximal conductance of e^2/h (1.0 on colour scale) along their traces in the range shown here. The base temperature is $T = 50$ mK; a perpendicular magnetic field of 6 T is applied to select a single spin species. White horizontal lines and 'X' are explained below. **b**, **c**, Resonant conductance for symmetric (**b**) and asymmetric (**c**) coupling, measured at several temperatures on the peak marked 'X' in **a**, as a function of ΔV_{gate}, the gate voltage relative to the centre of the peak. The side-gate voltages are fixed at the values indicated by white lines in **a**. As the temperature is reduced, in the symmetric case the peak becomes taller and narrower. By contrast, in the asymmetric case, the peak becomes shorter and its width saturates.

modes represented by a dynamical phase associated with the tunnelling matrix element[18]. We show that the analogy between tunnelling with dissipation and tunnelling in a Luttinger liquid[22,23,25,26] further extends to our case of resonant tunnelling. Based on our mapping and the Luttinger-liquid predictions[8–12], in the case of symmetric coupling we expect the peak height to saturate at e^2/h (spinless case), and the resonance width to scale at low temperatures as $T^{r/(r+1)}$ (ref. 8). For asymmetric coupling, the resonance width is predicted to saturate at sufficiently low temperature[10,11], while the peak height should scale to zero as $G \propto T^{2r}$, featuring the same exponent as in the single-barrier (non-resonant) tunnelling.

Our experiment clearly corroborates these predictions (Fig. 3). Quantitatively, we extract $r \approx 0.75$ from the scaling of the asymmetric peak height, which agrees with the leads resistance in this sample. The width of the symmetric peak scales with an exponent of 0.45, consistent with $r/(r+1) \approx 0.43$. (We discuss the accuracy of extracting the

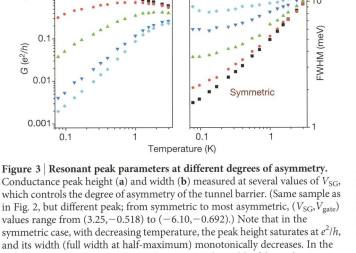


Figure 3 | Resonant peak parameters at different degrees of asymmetry. Conductance peak height (**a**) and width (**b**) measured at several values of V_{SG}, which controls the degree of asymmetry of the tunnel barrier. (Same sample as in Fig. 2, but different peak; from symmetric to most asymmetric, (V_{SG},V_{gate}) values range from (3.25,−0.518) to (−6.10,−0.692).) Note that in the symmetric case, with decreasing temperature, the peak height saturates at e^2/h, and its width (full width at half-maximum) monotonically decreases. In the asymmetric cases, the behaviour is the opposite: the width of the peak saturates, while the peak height decreases.

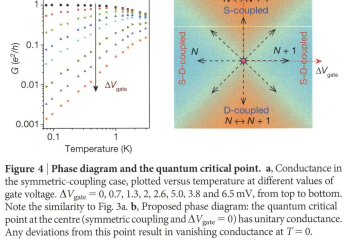


Figure 4 | Phase diagram and the quantum critical point. a, Conductance in the symmetric-coupling case, plotted versus temperature at different values of gate voltage. $\Delta V_{gate} = 0$, 0.7, 1.3, 2, 2.6, 5.0, 3.8 and 6.5 mV, from top to bottom. Note the similarity to Fig. 3a. **b**, Proposed phase diagram: the quantum critical point at the centre (symmetric coupling and $\Delta V_{gate} = 0$) has unitary conductance. Any deviations from this point result in vanishing conductance at $T = 0$.

exponents in Supplementary Information section III.) Overall, application of refs 8–12 to our experiment describes the observed behaviour remarkably well, in both the symmetric and asymmetric cases.

Note that the width of the conductance peak in the symmetric case decreases monotonically with decreasing temperature. In the limit of zero temperature, we expect that the conductance will equal zero everywhere, except for a singular point at the centre of the peak. When tunnelling asymmetry is introduced, the singular point disappears, and the low-temperature conductance tends to zero at any gate voltage, V_{gate}. This behaviour indicates a QPT for symmetric coupling, $\Gamma_S = \Gamma_D$. Technically, we refer to a 'boundary' QPT, in which only a local part of a larger system (local site plus environment) undergoes the transition (see, for example, ref. 5). QPTs found in strongly correlated bulk materials are often explained by invoking interactions between local sites and collective modes[4]. In the same spirit, our observation provides an example of a QPT in a highly tunable system that emulates such a 'local site' (that is, the resonant level), embedded in an interacting host.

Following ref. 12, we map our model in the $r = 1$ case onto the exotic two-channel Kondo model[27,28], for which a QPT is known to occur exactly for symmetric coupling[9]. (See Supplementary Information section V for details.) In both models, the origin of the quantum critical behaviour is the competition between the two channels attempting to screen the local site (spin or resonant level).

Intermediate values of r, between 0 and 1, do not allow for a simple interpretation in terms of any Kondo model with non-interacting leads, but represent a continuous evolution between the non-interacting resonant level at $r = 0$ and the two-channel Kondo model[9,29]. The critical exponents describing the system parameters close to the quantum critical point are not fixed, but are controlled by the value of r. Thus, our system not only provides new insight into the two-channel Kondo model—a model example of quantum criticality—but also gives access to a new family of quantum critical points for $r > 0$.

The QPT observed here is different from the various QPTs observed[30] and predicted[25] in quantum dots coupled to a single screening channel; indeed, there the QPTs are of the Kosterlitz–Thouless type, whereas in our case the QPT is of second order (see Supplementary Information section VI). Furthermore, in our case, the key ingredient that enables the QPT is the symmetric coupling to the two leads, which allows for their competition; the interaction in the leads (finite r) prevents their hybridization.

Finally, the conductance in the symmetric case can be plotted as a function of temperature for several values of ΔV_{gate} (Fig. 4a). The similarity with Fig. 3a is striking; apparently, one can tune away from the unitary resonance either by inducing asymmetry (Fig. 3a) or by applying the gate voltage (Fig. 4a)[8], with virtually the same results. Note that the downturn of peak height in Figs 3a and 4a occurs at progressively lower temperature as either the degree of asymmetry or ΔV_{gate} is reduced. Clearly, a new energy scale is emerging in the system, controlled by proximity to the quantum critical point[8,10,11]. We anticipate that this scale should vanish exactly at that point. We therefore propose the phase diagram shown in Fig. 4b, with a quantum critical point at $\Gamma_S = \Gamma_D$, $\Delta V_{gate} = 0$ (I. Affleck, personal communication). The four quadrants represent the states of the nanotube filled with N or $N + 1$ electrons, or coupled more strongly to either the source or the drain. The boundaries between the quadrants are smeared, and at $T = 0$ the conductance tends to zero everywhere, except at the quantum critical point.

In conclusion, we have investigated resonant tunnelling between interacting leads emulating Luttinger liquids. For symmetric coupling of the spinless resonant level to the two leads, and on resonance, the low-temperature conductance saturates at the unitary value of e^2/h. We associate this behaviour with a quantum critical point, which exists at $\Gamma_S = \Gamma_D$ in the presence of a finite interaction strength $r > 0$. Moving away from this point by inducing tunnelling asymmetry results in suppression of conductance at low temperature and smearing of the QPT. We believe that our work is the first example of a QPT in a highly tunable system, in which many parameters can be controlled, including the strength of interactions.

Received 13 April; accepted 24 May 2012.

1. Giamarchi, T. *Quantum Physics in One Dimension* (Oxford Univ. Press, 2004).
2. Chang, A. Chiral Luttinger liquids at the fractional quantum Hall edge. *Rev. Mod. Phys.* **75,** 1449–1505 (2003).
3. Deshpande, V. V., Bockrath, M. W., Glazman, L. I. & Yacoby, A. Electron liquids and solids in one dimension. *Nature* **464,** 209–216 (2010).
4. Sachdev, S. *Quantum Phase Transitions* 2nd edn (Cambridge Univ. Press, 2011).
5. Vojta, M. Impurity quantum phase transitions. *Phil. Mag.* **86,** 1807–1846 (2006).
6. Bloch, I. Ultracold quantum gases in optical lattices. *Nature Phys.* **1,** 23–30 (2005).
7. Si, Q. & Steglich, F. Heavy fermions and quantum phase transitions. *Science* **329,** 1161–1166 (2010).
8. Kane, C. L. & Fisher, M. P. A. Transmission through barriers and resonant tunnelling in an interacting one-dimensional electron gas. *Phys. Rev. B* **46,** 15233–15262 (1992).
9. Eggert, S. & Affleck, I. Magnetic impurities in half-integer-spin Heisenberg antiferromagnetic chains. *Phys. Rev. B* **46,** 10866–10883 (1992).
10. Nazarov, YuV & Glazman, L. I. Resonant tunnelling of interacting electrons in a one-dimensional wire. *Phys. Rev. Lett.* **91,** 126804 (2003).
11. Polyakov, D. G. & Gornyi, I. V. Transport of interacting electrons through a double barrier in quantum wires. *Phys. Rev. B* **68,** 035421 (2003).
12. Komnik, A. & Gogolin, A. O. Resonant tunnelling between Luttinger liquids: a solvable case. *Phys. Rev. Lett.* **90,** 246403 (2003).

13. Milliken, F., Umbach, C. & Webb, R. Indications of a Luttinger liquid in the fractional quantum Hall regime. *Solid State Commun.* **97,** 309–313 (1996).
14. Maasilta, I. & Goldman, V. Line shape of resonant tunnelling between fractional quantum Hall edges. *Phys. Rev. B* **55,** 4081–4084 (1997).
15. Kastner, M. A. Artificial atoms. *Phys. Today* **46,** 24–31 (1993).
16. Kouwenhoven, L. P. *et al.* in *Mesoscopic Electron Transport* (eds Sohn, L. L., Kouwenhoven, L. P. & Schön, G.) 105–214 (Kluwer, 1997).
17. Leggett, A. J. *et al.* Dynamics of the dissipative two-state system. *Rev. Mod. Phys.* **59,** 1–85 (1987).
18. Ingold, G.-L. & Nazarov, Y. V. in *Single Charge Tunnelling: Coulomb Blockade Phenomena in Nanostructures* (eds Grabert, H. & Devoret, M. H.) 21–107 (Plenum Press, 1992).
19. Flensberg, K., Girvin, S., Jonson, M., Penn, D. R. & Stiles, M. D. Quantum mechanics of the electromagnetic environment in the single-junction Coulomb blockade. *Physica Scripta* T42, 189–206 (1992). ∞.
20. Joyez, P., Esteve, D. & Devoret, M. H. How is the Coulomb blockade suppressed in high- conductance tunnel junctions? *Phys. Rev. Lett.* **80,** 1956–1959 (1998).
21. Zheng, W., Friedman, J., Averin, D. V., Han, S. & Lukens, J. E. Observation of strong Coulomb blockade in resistively isolated tunnel junctions. *Solid State Commun.* **108,** 839–843 (1998).
22. Sassetti, M., Napoli, F. & Weiss, U. Coherent transport of charge through a double barrier in a Luttinger liquid. *Phys. Rev. B* **52,** 11213–11224 (1995).
23. Safi, I&. Saleur, H. One-channel conductor in an ohmic environment: mapping to a Tomonaga-Luttinger liquid and full counting statistics. *Phys. Rev. Lett.* **93,** 126602 (2004).
24. Bomze, Y., Mebrahtu, H., Borzenets, I., Makarovski, A. & Finkelstein, G. Resonant tunnelling in a dissipative environment. *Phys. Rev. B* **79,** 241402(R) (2009).
25. Le Hur, K. & Li, M.-R. Unification of electromagnetic noise and Luttinger liquid via a quantum dot. *Phys. Rev. B* **72,** 073305 (2005).
26. Florens, S., Simon, P., Andergassen, S. & Feinberg, D. Interplay of electromagnetic noise and Kondo effect in quantum dots. *Phys. Rev. B* **75,** 155321 (2007).
27. Hewson, A. *The Kondo Problem to Heavy Fermions* (Cambridge Univ. Press, 1997).
28. Potok, R. M., Rau, I. G., Shtrikman, H., Oreg, Y. & Goldhaber-Gordon, D. J. Observation of the two-channel Kondo effect. *Nature* **446,** 167–171 (2007).
29. Goldstein, M. & Berkovits, R. Capacitance of a resonant level coupled to Luttinger liquids. *Phys. Rev. B* **82,** 161307 (2010).
30. Roch, N., Florens, S., Bouchiat, V., Wernsdorfer, W. & Balestro, F. Quantum phase transition in a single-molecule quantum dot. *Nature* **453,** 633–637 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We appreciate discussions with I. Affleck, D. V. Averin, A. M. Chang, C. H. Chung, S. Florens, M. Goldstein, L. I. Glazman, K. Ingersent, K. Le Hur, M. Lavagna, A. H. MacDonald, Yu. V. Nazarov, D. G. Polyakov and M. Vojta. We thank J. Liu for providing the nanotube growth facilities and W. Zhou for helping to optimize the nanotube synthesis. The work was supported by US DOE awards DE-SC0002765, DE-SC0005237 and DE-FG02-02ER15354.

Author Contributions H.T.M., I.V.B. and G.F. designed the experiment. H.T.M. fabricated the samples. H.T.M., I.V.B., Y.V.B., A.S. and G.F. conducted the experiment. H.T.M. and G.F. analysed the data. H.T.M, D.E.L., H.Z., H.U.B. and G.F. interpreted the data. D.E.L., H.Z. and H.U.B. developed the theory.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to G.F. (gleb@phy.duke.edu).

LETTER

doi:10.1038/nature11297

A Newtonian approach to extraordinarily strong negative refraction

Hosang Yoon[1], Kitty Y. M. Yeung[1], Vladimir Umansky[2] & Donhee Ham[1]

Metamaterials with negative refractive indices can manipulate electromagnetic waves in unusual ways, and can be used to achieve, for example, sub-diffraction-limit focusing[1], the bending of light in the 'wrong' direction[2], and reversed Doppler and Cerenkov effects[2]. These counterintuitive and technologically useful behaviours have spurred considerable efforts to synthesize a broad array of negative-index metamaterials with engineered electric, magnetic or optical properties[1–10]. Here we demonstrate another route to negative refraction by exploiting the inertia of electrons in semiconductor two-dimensional electron gases, collectively accelerated by electromagnetic waves according to Newton's second law of motion, where this acceleration effect manifests as kinetic inductance[11,12]. Using kinetic inductance to attain negative refraction was theoretically proposed for three-dimensional metallic nanoparticles[13] and seen experimentally with surface plasmons on the surface of a three-dimensional metal[14]. The two-dimensional electron gas that we use at cryogenic temperatures has a larger kinetic inductance than three-dimensional metals, leading to extraordinarily strong negative refraction at gigahertz frequencies, with an index as large as −700. This pronounced negative refractive index and the corresponding reduction in the effective wavelength opens a path to miniaturization in the science and technology of negative refraction.

The idea of creating negative refraction by exploiting the collective electron acceleration (inertia) effect, or kinetic inductance, was theoretically proposed for specific arrangements of three-dimensional (3D) metallic nanoparticles[13]. Experimentally, inertia-based negative refraction was implied in work where a particular guiding of surface plasmon polaritons on the surface of a 3D metal led to negative refraction[14]; this cannot be explained without electron acceleration, because a defining component of plasmons is the time-varying kinetic energy of their constituent electrons, which implies their acceleration.

Semiconductor two-dimensional (2D) electron gases (2DEGs) possess a much larger kinetic inductance than 3D bulk metals. Here we create negative-index metamaterials by fully exploiting this large kinetic inductance, whose impact is manifested in the extraordinarily large negative index, which we measure to be as large as $n = -700$. This is two orders of magnitude larger than the index of $n \approx -5$ to -1 in surface-plasmon-based negative refraction[14], and indicates that inertia is a much more important factor in our 2DEG case. It is also much larger than the theoretical expectation based on the kinetic inductance of 3D metallic nanoparticles[13], which is orders of magnitude smaller than our 2D kinetic inductance (Supplementary Information, section 1).

We choose a GaAs/AlGaAs 2DEG as a demonstration platform. Here electrons can accelerate for ~0.2 ns at a temperature of 4 K without scattering, with the result that their large kinetic inductance effect is not masked by the scattering at and above gigahertz frequencies. Specifically, our metamaterial is a periodic array of mesa-etched 2DEG strips (Fig. 1a, b), each of which is connected to ground lines (labelled 'G' in Fig. 1a) at both ends via ohmic contacts. Each strip's width and length are respectively denoted W and l, and the centre-to-centre distance between neighbouring strips, or periodicity, is denoted a. This metamaterial is excited by electromagnetic waves guided by the left signal line (labelled 'S' in Fig. 1a), which, flanked by the ground lines, forms an on-chip coplanar waveguide (CPW) with a 50-Ω characteristic impedance. This left signal line is extended to cover a few 2DEG strips on the left-hand side of the metamaterial, with dielectric between the signal line and the 2DEG strips. The metamaterial's response is picked up by the right signal line (also labelled 'S') of another CPW on the right-hand side of the metamaterial.

The electric fields of the excitation electromagnetic wave, oscillating between the signal and ground lines of the left CPW, collectively accelerate electrons in the leftmost few 2DEG strips, producing currents along the strips. The resulting alterations of charge distribution in these strips will capacitively couple to neighbouring strips to the right, accelerating electrons there. This process repeats to deliver an 'effective wave' from left to right, perpendicular to the direction of the strips. From the circuit point of view, each 2DEG strip—along which electrons collectively accelerate, with the resulting current lagging the accelerating voltage by 90° according to Newton's second law of motion—acts as non-magnetic inductance of kinetic origin[11,12]. This 2D kinetic inductance, $L_{k,2D}$, results from Newton's law: $L_{k,2D} = m^*/(n_{2D}e^2) \times (l_e/W)$, where m^*, e and n_{2D} are respectively the electrons' effective mass, charge and density per unit area, and l_e, which will be identified shortly, is the effective length of each strip, within which electrons accelerate in response to the excitation. Our metamaterial is then an array of capacitively coupled kinetic inductors (Fig. 1c and Supplementary Information, section 2), and may be

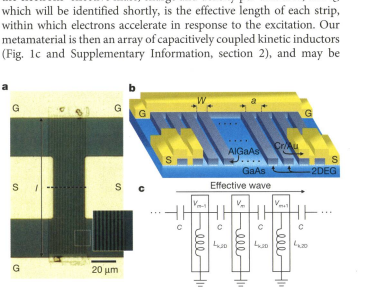


Figure 1 | Device description. a, Optical image of a 2DEG strip-array metamaterial prototype. Ground–signal–ground (GSG) on-chip CPWs direct electromagnetic waves to and from the metamaterial. The inset shows a magnified portion of the strip array. In this specific prototype, $W = 1$ μm, $l = 112$ μm and $a = 1.25$ μm. **b**, Schematic of the metamaterial (not drawn to scale), with the front face corresponding to a cut through the dashed symmetry line in **a**. **c**, Circuit description of the half of the metamaterial below or above the symmetry line along the effective wave propagation direction (Supplementary Information, section 2).

[1]School of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts 02138, USA. [2]Department of Condensed Matter Physics, Weizmann Institute of Science, Rehovot 76100, Israel.

likened to a left-handed transmission line[15–17], which is an array of capacitively coupled magnetic inductors and is known to be negatively refracting. However, our negative refraction originates in a different physical phenomenon: our device uses extremely large 2DEG kinetic inductance, whereas the left-handed transmission line relies on a much smaller magnetic inductance.

To examine the negative refraction behaviour of our device, we represent the effective wave, in terms of the voltage at the tip of the mth kinetic inductor (Fig. 1c), as $V_m(t) \propto e^{i(\omega t - mka)}$, where ω is the angular frequency and k is the effective wavenumber. The standard circuit analysis of Fig. 1c yields a dispersion relation $\omega(k) = \omega_c|\sin(ka/2)|$, where $\omega_c \equiv [2\sqrt{(L_{k,2D}C)}]^{-1}$ is the cut-off frequency at the boundary of the first Brillouin zone ($k = \pm\pi/a$) and C is the capacitance between adjacent strips over the effective length (Supplementary Information, section 2). For $\omega > \omega_c$, the dispersion relation (Fig. 2a) predicts negative refraction, because the tangential slope $d\omega/dk$ (the group velocity) and the slope ω/k (the phase velocity) have opposite signs[10]. The cut-off behaviour results from the metamaterial's high-pass nature, and can also be seen from the current distributions across the metamaterial below and above the cut-off frequency (Fig. 2b), which we simulated using an electromagnetic field solver (Supplementary Information, section 6). Beyond the cut-off frequency (Fig. 2b, right), the current is concentrated at the bottom and top regions of the strips, from which l_e can be estimated. We note that, whereas a single sheet of 2DEG exhibits ordinary dispersion[18], the acceleration of electrons along an array of strips of 2DEG, perpendicular to the direction of effective wave propagation, causes negative refraction.

The dispersion relation of our metamaterial has the same form as that of the left-handed transmission line[15–17], but with the magnetic inductance replaced with the much larger 2DEG kinetic inductance; the 2DEG kinetic inductance is 1.25 nH μm^{-1} for a 1-μm wide 2DEG strip, which is ~2,800 times larger than the same strip's magnetic inductance, 0.44 pH μm^{-1} (Supplementary Information, section 1). The effective refractive index derived from the dispersion is $n = -2c/(a\omega) \times \sin^{-1}(\omega_c/\omega)$, where c is the speed of light in vacuum, and has the maximum attainable magnitude of $2c/(a\omega_c) = 4c/a \times \sqrt{(L_{k,2D}C)}$, which is exceedingly large owing to the large 2DEG kinetic inductance, corresponding to the substantial slowing of the effective wave.

Microwave scattering experiments with on-chip probing confirm this extraordinarily strong negative refraction. The reflection of an electromagnetic wave incident on the left on-chip CPW and its transmission to the right on-chip CPW after propagation through the metamaterial are measured over a range of ~1–50 GHz using a vector network analyser. Propagation delays in the two on-chip CPWs and parasitic couplings between them bypassing the metamaterial were separately measured and de-embedded; from the resulting transmission and reflection coefficients, s_{21} and s_{11}, at each measurement frequency, we extract, using a well-established method[19–23], the effective wave's phasor change e^{-ikd} due purely to propagation of a distance d across the metamaterial (Supplementary Information, section 3).

Figure 3a shows the frequency–wavenumber (f–k) dispersion so obtained at temperatures of 4.2, 10 and 20 K for a 13-strip metamaterial with $W = 1$ μm, $l = 112$ μm, $l_e = 31$ μm and $a = 1.25$ μm. Because the measured parameters s_{22} and s_{11} set the left-to-right energy propagation direction (that is, the direction of the group velocity) as the positive reference direction, if our metamaterial is negatively refracting, the sign of the extracted wavenumber will be negative with no ambiguity, which is indeed seen in Fig. 3a. Negative refraction is also consistently confirmed in Fig. 3a by the fact that $d\omega/dk$ and ω/k have opposite signs above the 12-GHz cut-off frequency. This measured dispersion, including the cut-off frequency, differs in its details from the calculation that uses lumped circuit elements, ignores losses due to electron scattering in the 2DEG strips and ohmic contacts, and considers only nearest capacitive couplings (Fig. 2). But it has the same underlying features, demonstrating negative refraction. The dark area in Fig. 3a, where the distinctively spurious behaviour of the dispersion appears, is indicative of the cut-off region, which is irrelevant to the operation of the device (Supplementary Information, section 3). From this f–k dispersion, we obtain the effective refractive index using $n = kc/\omega$, whose real part is as large as −500 (Fig. 3b). This large negative index, which is difficult, if not impossible, to achieve with magnetic inductance[3–5,15–17,24,25], allows drastic device miniaturization and can facilitate ultra-subwavelength localization. The same measurements performed on the 2DEG strip array, but with energy propagation along the strips, yield positive refraction, further highlighting our negative refraction strategy (Supplementary Information, section 5).

We find that $|\mathrm{Re}(n)|$ decreases with frequency (Fig. 3b), because as the frequency increases adjacent strips are coupled more capacitively, which increasingly bypasses the electron acceleration effect within each separated strip. Figure 3c shows the figure of merit, $|\mathrm{Re}(n)/\mathrm{Im}(n)|$, which here reflects losses due to electron scattering in the 2DEG strips and ohmic contacts. It takes a value of ~2 over a reasonably large part of the negative refraction region, similarly to negative refraction devices using metals at optical frequencies[24,25]. Figure 3a–c shows that the negative refraction behaviour is essentially the same regardless of temperature (4.2, 10 and 20 K), indicating that the degree of electron scattering in the 2DEG strips and ohmic contacts remains largely the same within this temperature range, not masking the inertia effect. In Fig. 3c, the figure of merit is largest at 10 K instead of at 4.2 K, but these variations with respect to temperature arise mostly from inconsistent probe landings during multiple calibration steps, which are done for measurements at each temperature. Fluctuations at high frequencies, for example, in Fig. 3c, are also due to imperfect calibration.

At 297 K, electron scattering in each 2DEG strip becomes so severe that the acceleration effect is completely masked. Equivalently, the strip's ohmic resistance becomes far larger (~100 kΩ) than the impedance of its kinetic inductance. The strip array then becomes essentially an open circuit, causing the signal to be mostly reflected. This reflection can be seen from the value of the reflection coefficient, $|s_{11}| \approx 1$, measured at 297 K, which differs from $|s_{11}|$ at cryogenic temperatures, where the strip array exhibits negative refraction (Fig. 3d). The transmission coefficient, $|s_{21}|$, becomes smaller at 297 K also because of the open-circuit behaviour, but is not unappreciable (Fig. 3d). To understand this, we fabricated exactly the same structure as the previous device, but without the strip array, thus creating an actual open circuit between the two on-chip CPWs. The behaviour of $|s_{21}|$ for this open device at 4.2 K closely resembles that of $|s_{21}|$ for the strip array at 297 K (Fig. 3d). This demonstrates that the behaviour in the strip array at 297 K is due largely to the parasitic coupling between the two CPWs bypassing the

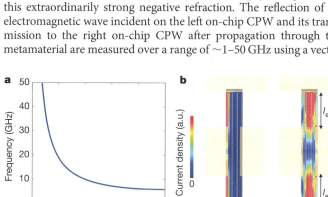


Figure 2 | Theory and simulation. **a**, Plot of $\omega(k) = [2\sqrt{(L_{k,2D}C)}|\sin(ka/2)|]^{-1}$, with $L_{k,2D} = 39$ nH and $C = 4.6$ fF estimated for the structure measured for Fig. 3 ($W = 1$ μm, $a = 1.25$ μm, $l = 112$ μm). The group and phase velocities, $d\omega/dk$ and ω/k, have opposite signs, showing negative refraction; this occurs for both $k > 0$ and $k < 0$, but we show only the latter, which is relevant to our measurements. **b**, Simulated current distributions below (left; 5 GHz) and above (right; 30 GHz) the cut-off frequency. Red and blue colours indicate high and low current densities, respectively. Above the cut-off frequency, regions of high, constant current density are observed, from which the effective strip length l_e is estimated. a.u., arbitrary units.

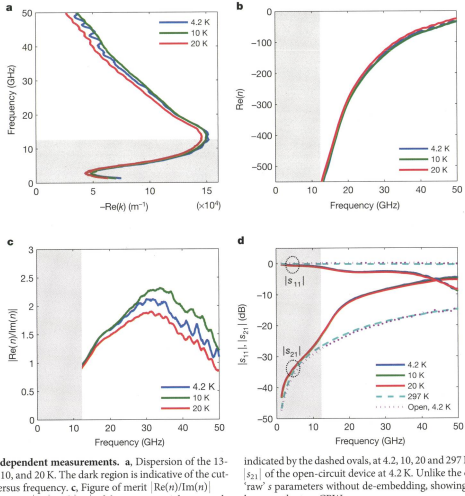


Figure 3 | Temperature-dependent measurements. **a**, Dispersion of the 13-strip metamaterial at 4.2, 10, and 20 K. The dark region is indicative of the cut-off behaviour. **b**, Re(n) versus frequency. **c**, Figure of merit $|\mathrm{Re}(n)/\mathrm{Im}(n)|$ versus frequency. **d**, Parameters $|s_{11}|$ and $|s_{21}|$ of the metamaterial, separately indicated by the dashed ovals, at 4.2, 10, 20 and 297 K. Also shown are $|s_{11}|$ and $|s_{21}|$ of the open-circuit device at 4.2 K. Unlike the data in **a**, **b** and **c**, these are 'raw' s parameters without de-embedding, showing the parasitic coupling between the two CPWs.

strip array, confirming the open-circuit nature of the device at 297 K (in fact, the phases of the transmission and reflection parameters are also much the same in the open device at 4.2 K and the strip array at 297 K; see Supplementary Information, section 4). These results provide further confirmation that the negative refraction we observe at cryogenic temperatures is due to the kinetic inductance. Also, in most of the dark area in Fig. 3d, the $|s_{21}|$ value of the metamaterial even at cryogenic temperatures is very similar to that of the open device, confirming the cut-off nature in that region.

To examine further the impact of kinetic inductance on negative refraction, we measure a new set of devices of various geometric parameters. Comparison of devices with different values of l (and, thus, l_k) for the same values of W and a is especially instructive; changing l scales $L_{k,2D}$, C and ω_c proportionally to l_e, affecting the index $n = -(2c/a\omega) \times \sin^{-1}(\omega_c/\omega)$ with only one parameter, ω_c. Specifically, a device with longer strips, with larger values of $L_{k,2D}$ and C, and a smaller ω_c value, will have a negative index with a larger maximum attainable magnitude, $2c/a\omega_c$, reaching the frequency region forbidden for a shorter-strip device. In the frequency region accessible by both longer- and shorter-strip devices, the shorter-strip device will have a larger negative index than the longer-strip device if the two are compared at the same frequency (Supplementary Fig. 4). This clear-cut property emerges in measurements of a pair of devices both with $a = 1.25$ μm but with differing values of l (112 versus 52 μm) or l_e (31 versus 14 μm) (Fig. 4a). This property is confirmed again in two additional pairs of devices (Fig. 4b), where the refractive index is as large as -700.

Altering the periodicity a, which in general may have to be combined with altering W, affects the index $n = -(2c/a\omega) \times \sin^{-1}(\omega_c/\omega)$ in a more complicated manner, owing to simultaneous changes in a and ω_c. For $l = 112$ μm, as we decrease (a, W) first from (1.5 μm, 1 μm) to (1.25 μm, 1 μm) and then to (0.75 μm, 0.6 μm), with the first reduction increasing C by a factor of 1.2 with $L_{k,2D}$ unchanged, and the second reduction increasing $L_{k,2D}$ by a factor of 1.7 with C unchanged, ω_c does not vary as much as a, owing to the square-root dependence of ω_c on $L_{k,2D}$ and C. Thus, a smaller periodicity will yield a larger negative index for the same frequency away from the cut-off regions, as evident in measurements (Fig. 4c). In these a and W variations, the characteristic impedance and, thus, the impedance mismatch is varied. This results in imperfect de-embedding of the parasitic couplings near cut-offs, obscuring the cut-off behaviours. The tendency of the index to be more negative for smaller periodicities is seen again for $l = 52$ μm with the same variations of a and W (Fig. 4d). The crossing of the respective data for the devices with $a = 1.25$ and 1.5 μm is an anomaly that we suspect arises from the impedance mismatch variation.

The exceedingly strong inertia-based negative refraction demonstrated here requires a solid-state platform with a very large kinetic inductance and low electron scattering. To meet these requirements, we used a GaAs/AlGaAs 2DEG at cryogenic temperature. Scaling the 2DEG metamaterial to higher frequencies by simultaneous reduction of the strip length and periodicity (Fig. 4 and Supplementary Information, section 2) would relax the condition on electron scattering time (and, thus, temperature); demonstration[26] of terahertz plasmonic devices at room temperature with GaAs/AlGaAs 2DEG bodes well for high-temperature applications. Graphene, another type of 2D conductor with high mobility at room temperature[27], may also be a platform for terahertz-frequency negative refraction based on a

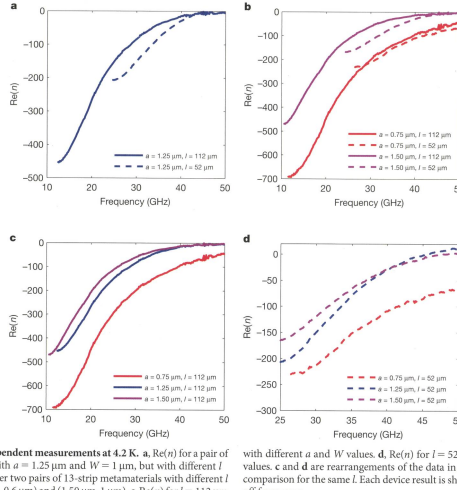


Figure 4 | Geometry-dependent measurements at 4.2 K. a, Re(n) for a pair of 13-strip metamaterials with $a = 1.25$ μm and $W = 1$ μm, but with different l values. **b**, Re(n) for another two pairs of 13-strip metamaterials with different l values: (a, W) = (0.75 μm, 0.6 μm) and (1.50 μm, 1 μm). **c**, Re(n) for $l = 112$ μm with different a and W values. **d**, Re(n) for $l = 52$ μm with different a and W values. **c** and **d** are rearrangements of the data in **a** and **b** to facilitate comparison for the same l. Each device result is shown above its respective cut-off frequency.

similar kinetic approach. Although electrons in graphene act as massless particles, and thus are non-Newtonian, they still possess kinetic energy, exhibiting plasmonic behaviour with implicit kinetic inductance. In fact, terahertz light–plasmon coupling has been recently observed at room temperature[28]. Achieving strong negative refraction based on the kinetic approach with higher isotropy and at optical frequencies using different material systems is also open to further investigation.

METHODS SUMMARY

We fabricate the devices on GaAs/AlGaAs 2DEG substrates obtained by molecular beam epitaxy. The layer structure above the 2DEG comprises 40-nm $Al_{0.36}Ga_{0.64}As$, 14-nm Si-doped $Al_{0.36}Ga_{0.64}As$, 10-nm $Al_{0.36}Ga_{0.64}As$ and a 7-nm GaAs cap. At 4 K, the mobility of the 2DEG is 4.6×10^6 cm^2 V^{-1}s^{-1} and the carrier density is 1.9×10^{11} cm^{-2}, both in the dark. 2DEG strips are defined by electron beam lithography, followed by wet etching (>71-nm depth) with 150:1:1 H_2O:H_2O_2:NH_4OH. Ohmic contacts are defined by photolithography followed by thermal evaporation of Ni (5 nm)/Au (20 nm)/Ge (25 nm)/Au (10 nm)/Ni (5 nm)/Au (40 nm), and annealing at 420 °C for 50 s. CPWs are defined by photolithography and formed by thermal evaporation of Cr (8 nm)/Au (500 nm).

The microwave scattering analysis is performed in a Lake Shore Cryotronics cryogenic probe station at feedback-controlled cryogenic temperatures in the dark. Ground–signal–ground microwave probes, with a pitch of 100 μm, connected to probe arms are attached to on-chip CPWs. Coaxial cables lead to the probes from an Agilent E8364A network analyser, which generates excitation signals of frequencies up to 50 GHz, delivering −45 dBm power to the devices, and measures the scattering parameters. To see the effect of the metamaterial (strip array) only, we first calibrate the system, at each measurement temperature, up to the tips of the probes by using the NIST-style multiline TRL technique[29], and then perform additional de-embedding to remove the on-chip CPW delays and parasitic couplings between the CPWs, which bypass the metamaterial (Supplementary Information, section 3). The CPWs used for this calibration[30] are fabricated on undoped GaAs substrates and designed using a Sonnet electromagnetic solver to have a 50-Ω characteristic impedance, which is the characteristic impedance of the network analyser, cables and probes.

Received 2 December 2011; accepted 31 May 2012.

1. Pendry, J. B. Negative refraction makes a perfect lens. *Phys. Rev. Lett.* **85,** 3966–3969 (2000).
2. Veselago, V. G. The electrodynamics of substances with simultaneously negative values of ε and μ. *Sov. Phys. Usp.* **10,** 509–514 (1968).
3. Smith, D. R., Padilla, W. J., Vier, D. C., Nemat-Nasser, S. C. & Schultz, S. Composite medium with simultaneously negative permeability and permittivity. *Phys. Rev. Lett.* **84,** 4184–4187 (2000).
4. Shelby, R. A., Smith, D. R. & Schultz, S. Experimental verification of a negative index of refraction. *Science* **292,** 77–79 (2001).
5. Linden, S. *et al.* Photonic metamaterials: magnetism at optical frequencies. *IEEE J. Sel. Top. Quantum Electron.* **12,** 1097–1105 (2006).
6. Cubukcu, E., Aydin, K., Ozbay, E., Foteinopoulou, S. & Soukoulis, C. M. Electromagnetic waves: negative refraction by photonic crystals. *Nature* **423,** 604–605 (2003).
7. Valentine, J. *et al.* Three-dimensional optical metamaterial with a negative refractive index. *Nature* **455,** 376–379 (2008).
8. Podolskiy, V. A. & Narimanov, E. E. Strongly anisotropic waveguide as a nonmagnetic left-handed system. *Phys. Rev. B* **71,** 201101 (2005).
9. Hoffman, A. J. *et al.* Negative refraction in semiconductor metamaterials. *Nature Mater.* **6,** 946–950 (2007).
10. Pendry, J. B. A chiral route to negative refraction. *Science* **306,** 1353–1355 (2004).
11. Meservey, R. Measurements of the kinetic inductance of superconducting linear structures. *J. Appl. Phys.* **40,** 2028–2034 (1969).
12. Burke, P. J., Spielman, I. B., Eisenstein, J. P., Pfeiffer, L. N. & West, K. W. High frequency conductivity of the high-mobility two-dimensional electron gas. *Appl. Phys. Lett.* **76,** 745–747 (2000).
13. Engheta, N. Circuits with light at nanoscales: optical nanocircuits inspired by metamaterials. *Science* **317,** 1698–1702 (2007).
14. Lezec, H. J., Dionne, J. A. & Atwater, H. A. Negative refraction at visible frequencies. *Science* **316,** 430–432 (2007).

15. Eleftheriades, G. V., Iyer, A. K. & Kremer, P. C. Planar negative refractive index media using periodically L–C loaded transmission lines. *IEEE Trans. Microw. Theory Tech.* **50,** 2702–2712 (2002).
16. Caloz, C. & Itoh, T. Transmission line approach of left-handed (LH) materials and microstrip implementation of an artificial LH transmission line. *IEEE Trans. Antenn. Propag.* **52,** 1159–1166 (2004).
17. Grbic, A. & Eleftheriades, G. V. Overcoming the diffraction limit with a planar left-handed transmission-line lens. *Phys. Rev. Lett.* **92,** 117403 (2004).
18. Stern, F. Polarizability of a two-dimensional electron gas. *Phys. Rev. Lett.* **18,** 546–548 (1967).
19. Chen, X., Grzegorczyk, T. M., Wu, B.-I., Pacheco, J. & Kong, J. A. Robust method to retrieve the constitutive effective parameters of metamaterials. *Phys. Rev. E* **70,** 016608 (2004).
20. Smith, D. R., Vier, D. C., Koschny, T. & Soukoulis, C. M. Electromagnetic parameter retrieval from inhomogeneous metamaterials. *Phys. Rev. E* **71,** 036617 (2005).
21. Burgos, S. P., de Waele, R., Polman, A. & Atwater, H. A. A single-layer wide-angle negative-index metamaterial at visible frequencies. *Nature Mater.* **9,** 407–412 (2010).
22. Choi, M. *et al.* A terahertz metamaterial with unnaturally high refractive index. *Nature* **470,** 369–373 (2011).
23. Chanda, D. *et al.* Large-area flexible 3D optical negative index metamaterial formed by nanotransfer printing. *Nature Nanotechnol.* **6,** 402–407 (2011).
24. Shalaev, V. M. Optical negative-index metamaterials. *Nature Photon.* **1,** 41–48 (2007).
25. Soukoulis, C. M., Linden, S. & Wegener, M. Negative refractive index at optical wavelengths. *Science* **315,** 47–49 (2007).
26. Meziani, Y. M. *et al.* Room temperature terahertz emission from grating coupled two-dimensional plasmons. *Appl. Phys. Lett.* **92,** 201108 (2008).
27. Dean, C. R. *et al.* Boron nitride substrates for high-quality graphene electronics. *Nature Nanotechnol.* **5,** 722–726 (2010).
28. Ju, L. *et al.* Graphene plasmonics for tunable terahertz metamaterials. *Nature Nanotechnol.* **6,** 630–634 (2011).
29. Marks, R. B. A multiline method of network analyzer calibration. *IEEE Trans. Microw. Theory Tech.* **39,** 1205–1215 (1991).
30. Andress, W. F. *et al.* Ultra-subwavelength two-dimensional plasmonic circuits. *Nano Lett.* **12,** 2272–2277 (2012).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements The authors are grateful for support for this research by the Air Force Office of Scientific Research under contract numbers FA 9550-09-1-0369 and FA 9550-08-1-0254. Device fabrication was performed in part at the Center for Nanoscale Systems at Harvard University. The authors thank W. F. Andress for assistance with device fabrication and microwave measurements.

Author Contributions H.Y. and D.H. had the idea for the project. V.U. fabricated the 2DEG. H.Y. designed, fabricated and measured the properties of the devices. H.Y., K.Y.M.Y. and D.H. analysed the data. H.Y. and D.H. wrote the paper. All authors discussed the results and reviewed the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to D.H. (donhee@seas.harvard.edu).

LETTER

doi:10.1038/nature11299

Increase in observed net carbon dioxide uptake by land and oceans during the past 50 years

A. P. Ballantyne[1]†, C. B. Alden[2], J. B. Miller[3,4], P. P. Tans[4] & J. W. C. White[1,2]

One of the greatest sources of uncertainty for future climate predictions is the response of the global carbon cycle to climate change[1]. Although approximately one-half of total CO_2 emissions is at present taken up by combined land and ocean carbon reservoirs[2], models predict a decline in future carbon uptake by these reservoirs, resulting in a positive carbon–climate feedback[3]. Several recent studies suggest that rates of carbon uptake by the land[4–6] and ocean[7–10] have remained constant or declined in recent decades. Other work, however, has called into question the reported decline[11–13]. Here we use global-scale atmospheric CO_2 measurements, CO_2 emission inventories and their full range of uncertainties to calculate changes in global CO_2 sources and sinks during the past 50 years. Our mass balance analysis shows that net global carbon uptake has increased significantly by about 0.05 billion tonnes of carbon per year and that global carbon uptake doubled, from 2.4 ± 0.8 to 5.0 ± 0.9 billion tonnes per year, between 1960 and 2010. Therefore, it is very unlikely that both land and ocean carbon sinks have decreased on a global scale. Since 1959, approximately 350 billion tonnes of carbon have been emitted by humans to the atmosphere, of which about 55 per cent has moved into the land and oceans. Thus, identifying the mechanisms and locations responsible for increasing global carbon uptake remains a critical challenge in constraining the modern global carbon budget and predicting future carbon–climate interactions.

Coupled climate/carbon-cycle models predict decreased carbon (C) uptake by the land, owing to diminishing productivity and increasing respiration, and decreased C uptake by the ocean, associated with acidification, changes in ocean mixing and increasing sea surface temperatures, within this century[3]. Although detecting changes in regional C sinks is very challenging, several recent studies suggest that C uptake by the land and ocean may already be tapering off or declining. However, diminished C uptake in these studies is often limited to the regional[5,7–9] or decadal scale[4,6,10]. In addition, trends in sink intensity in these studies are inferred from satellite measurements[6], simulated using models[8,10] or estimated on the basis of inventories of existing C sinks[4,7,9]. Thus, their implications for long-term variation in the global C budget remain uncertain. Here we focus strictly on global-scale observations provided by atmospheric CO_2 measurements and CO_2 emission estimates, and include the full range of uncertainties in each to estimate changes in global C uptake during the past 50 yr. Although this 'top-down' approach does not provide the detailed process-level information of previous studies, it does provide an unbiased assessment of changes in global C uptake.

The growth rate of atmospheric CO_2

$$\frac{dC}{dt} = \sum F + \sum N \quad (1)$$

varies in response to one-way fluxes to the atmosphere (ΣF) and net exchange between the Earth's surface reservoirs and atmosphere (ΣN). The one-way fluxes include those from fossil fuel emissions, including cement production (F_F), and those from land-use change (F_L). Negative ΣN values represent net uptake of CO_2 and comprise contributions by the land (N_L) and the oceans (N_O). From the observed atmospheric growth rate and estimated fluxes, we can calculate net global CO_2 uptake ($\Sigma N = dC/dt - \Sigma F$) and the airborne fraction ($AF = (dC/dt)/\Sigma F$). Although both F_F and F_L are often included in calculating AF and ΣN, it can be argued that only F_F should be included in these calculations because it represents the addition of truly extrinsic C to the modern C cycle, which will be redistributed between the atmosphere, oceans and land. We calculate two versions of ΣN and AF, one with F_F and F_L (Figs 1 and 2 and Table 1), and one with only F_F (Table 1).

A major difficulty in characterizing the uncertainty of trends in ΣN and AF is that dC/dt errors are negatively autocorrelated in successive years, whereas ΣF errors are positively autocorrelated. The uncertainty of a trend will be overestimated if the negative autocorrelation is not taken into account in the analysis, and will be underestimated if the positive autocorrelation is not taken into account. Thus, calculations of ΣN and AF contain both positive and negative autocorrelations, the balance of which is time dependent. Before 1980 uncertainties in ΣN and AF are dominated by dC/dt errors, whereas towards the end of the record uncertainties in ΣN and AF become increasingly dominated by ΣF. To take this error structure properly into account, we used a Monte-Carlo-type approach to simulate the errors. To evaluate the

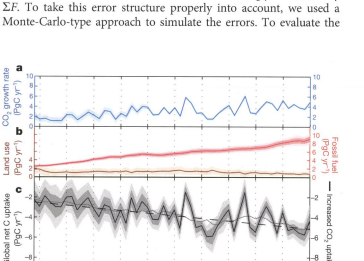


Figure 1 | Trends in the global carbon budget from 1959 to 2010. a, The annual atmospheric CO_2 growth rate (dC/dt). **b**, Fluxes of C to the atmosphere from fossil fuel emissions (F_F) are plotted in red and those from land-use changes (F_L) are plotted in brown. **c**, Annual global net C uptake (ΣN) is plotted as a black solid line and is compared with the 10-yr moving average (dark grey line) and the significant linear trend (dashed line) (Table 1). All dark shaded bands represent 1σ uncertainties and all light shaded bands represent 2σ uncertainties. Note that the scale of the y axis in **c** has been expanded.

[1]Department of Geology, University of Colorado, Boulder, Colorado 80309, USA. [2]Institute of Arctic and Alpine Research, University of Colorado, Boulder, Colorado 80309, USA. [3]Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, Colorado 80309, USA. [4]Earth System Research Laboratory, National Oceanographic and Atmospheric Administration, Boulder, Colorado 80305, USA. †Present address: Department of Ecosystem and Conservation Sciences, University of Montana, Missoula, Montana 59812, USA.

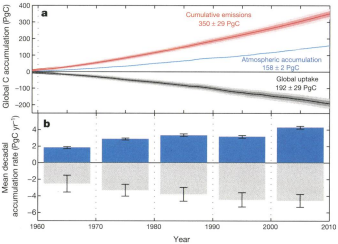


Figure 2 | Accumulation of carbon emissions in the atmosphere, on land and in the oceans. a, Sums of emissions from fossil fuels and land-use change integrated from 1959 to 2010 (red) are compared with atmospheric accumulation (blue) and cumulative global uptake (black) by the land and oceans. The dark shaded bands represent 1σ uncertainties and the light shaded bands represent 2σ uncertainties. **b**, Mean decadal C accumulation rates for the atmosphere (blue) and mean global C uptake rates (grey) are calculated as the sum of C accumulation over a given decade divided by 10 yr. Error bars represent the 2σ uncertainties (Methods).

uncertainty in dC/dt since 1980, atmospheric sampling sites were resampled and global mean growth rates were calculated to account for the spatial variability and sparse sampling of atmospheric CO_2, which contribute much more to uncertainty than do measurement and calibration errors. Before 1980, a negatively autocorrelated error component was added to dC/dt simulations. To assess the uncertainty in ΣF, we combined three independent inventories of F_F emissions[14–16] with three independent inventories of F_L emissions[17–19]. To each emission inventory, we added positively autocorrelated random errors to account for temporally persistent accounting errors[20]. These ΣF emission scenarios were then combined with the simulations of dC/dt to estimate trends in ΣN and AF (Methods).

Significantly increasing linear trends in observed dC/dt (0.054 ± 0.011 billion tonnes of carbon (PgC) per year per year) and estimated F_F (0.115 ± 0.011 PgC yr^{-2}) are evident, whereas F_L shows a slight decline between 1959 and 2010 (Fig. 1 and Table 1). Whereas the uncertainty in dC/dt has decreased over time, owing to the addition of network sites monitoring global atmospheric CO_2, the uncertainty in F_F has increased over time, primarily as a result of growing emissions and a greater contribution from emerging economies. We find a significant negative trend in ΣN of -0.052 ± 0.026 PgC yr^{-2} (Fig. 1c and Table 1), indicating a large increase in net global C uptake during the past 50 yr. Net global C uptake has grown on average by 0.5 PgC yr^{-1} per decade, from 2.4 ± 0.8 PgC yr^{-1} in 1960 to 5.0 ± 0.9 PgC yr^{-1} in 2010.

Superimposed on this increasing trend in net global C uptake is considerable variability (Fig. 1c). Although net global C uptake increased steadily from 1960 to about 1990, substantial oscillations in net global C uptake have occurred over the past 20 yr. In fact, an increasing trend in ΣN of 0.21 ± 0.10 PgC yr^{-2} was observed during the 1990s, but this was followed by an equally large decreasing trend in ΣN of -0.19 ± 0.08 PgC yr^{-2} since 2000. Thus, it might be inferred that global C sinks diminished during the 1990s; however, this apparent trend is mainly due to the timing of the eruption of Mt Pinatubo in 1991, which enhanced net global C uptake, and the strong El Niño event in 1998, which diminished net global C uptake[11].

A commonly used diagnostic for detecting changes in the relative C sink efficiency is the airborne fraction, AF. Our analysis reveals that trends in AF are highly sensitive to whether land-use emissions are included in the global C budget. When only fossil fuel emissions are included, there is a significant decreasing trend in AF, indicating an increase in uptake efficiency. In contrast, when both land-use and fossil fuel emissions are included, the sign of the rate of change of AF switches and the uncertainty range is larger, including both positive and negative trends (Table 1). There has been considerable debate as to whether AF has changed over time and what changes in AF indicate[12,13,21,22]. Our results show that when land-use emissions are included, there is no detectable change in AF over the last 50 yr. Our findings are corroborated by a recent independent analysis showing no significant change in AF since 1850[12]. Moreover, it has been demonstrated that large changes in uptake efficiency are required to alter AF significantly[13]. Thus, changes in AF over time are highly sensitive to land-use emissions and are difficult to interpret, whereas the significant trend in ΣN provides unequivocal evidence that net global CO_2 uptake continues to increase.

Alternatively, we investigate where anthropogenic emissions have accumulated between 1959 and 2010. Approximately 60 PgC from land use and 290 PgC from fossil fuels have been emitted to the atmosphere, making a total of 350 ± 29 PgC of anthropogenic emissions during that time frame (Fig. 2a). Of these, 158 ± 2 PgC remain in the atmosphere and 192 ± 29 PgC have accumulated in combined land and ocean reservoirs. Thus, 55% of anthropogenic CO_2 emissions have been transferred to the land and oceans, and 45% have remained in the atmosphere. The mean decadal C accumulation rate (Fig. 2b) in land and oceans has increased every decade, from 2.5 ± 1.0 PgC yr^{-1} during the 1960s to 4.6 ± 0.7 PgC yr^{-1} since 2000. In the atmosphere, the average decadal accumulation has increased from 1.8 ± 0.12 PgC yr^{-1} during the 1960s to 4.1 ± 0.06 PgC yr^{-1} between 2000 and 2010. Although the 1990s seem to be anomalous, in that a much greater proportion of C accumulated on land and in the oceans than in the atmosphere, since 2000 the rate of accumulation in the atmosphere has accelerated.

Because the trend in ΣN shows an increase in net global C uptake, N_L and N_O cannot both be decreasing. If regional ocean and land C sinks are indeed diminishing[5,7–9], then to satisfy the global C mass balance, these reduced sinks must be more than compensated for by an increase in the rate of uptake by existing C sinks or the formation of new C sinks. A global inventory of C dynamics in established forests has identified strong regional differences in uptake but a fairly constant global average uptake rate of approximately 2.5 PgC yr^{-1} over the past

Table 1 | Trend analyses of parameters and diagnostics of the global C budget (equation (1)) from 1959 to 2010

Parameters and diagnostics	Slope trend* (PgC yr^{-2})	95% confidence interval (PgC yr^{-2})	
		Minimum	Maximum
dC/dt	**0.054**	**0.043**	**0.065**
F_F	**0.115**	**0.103**	**0.126**
F_L	−0.007	−0.041	0.027
AF with F_F only	**−0.0016**	**−0.0029**	**−0.0002**
AF with F_F and F_L	0.0012	−0.0008	0.0032
$\sum N$ with F_F only	**−0.063**	**−0.076**	**−0.051**
$\sum N$ with F_F and F_L	**−0.052**	**−0.077**	**−0.026**

The mean values of slope trend and their 95% confidence interval are calculated from the distribution of simulations (Methods). Significant trends are in bold.
*AF is dimensionless.

20 yr (ref. 4). Similarly, trends in the partial pressure of CO_2 in the ocean indicate a decrease in C uptake in the Atlantic and an increase in C uptake in the Pacific over the past 30 yr (ref. 10). Thus, evidence for a change in the rate of C uptake by existing regional C sinks on land and in the ocean is equivocal. It has been suggested that widespread drought in the Southern Hemisphere has led to a decrease in terrestrial CO_2 uptake[6] and that increased surface wind velocity has led to decreased CO_2 uptake in the Southern Ocean[8]. Unfortunately, the atmospheric CO_2 observations required to validate these reported declines in Southern Hemisphere CO_2 uptake remain scarce.

From a global mass balance perspective, net uptake of atmospheric CO_2 has continued to increase during the past 50 yr and seems to remain strong. Although present predictions indicate diminished C uptake by the land and oceans in the coming century, with potentially serious consequences for the global climate, as of 2010 there is no empirical evidence that C uptake has started to diminish on the global scale. Therefore, to improve our understanding of carbon–climate interactions, more process studies focusing on mechanisms and regions of increased net CO_2 uptake are required, uncertainty in the global C budget must be reduced by better constraining estimates of fossil fuel emissions, and the global network monitoring atmospheric CO_2 must be expanded to include regions where C uptake is sensitive to climate variability. A fully comprehensive and credible global carbon budget can be achieved only when regional process studies are confirmed by global-scale observations.

METHODS SUMMARY

We use a Monte Carlo approach to assess uncertainty because it permits us to simulate the time-dependent autocorrelation structure of the uncertainty. From 1980 to 2010, the global atmospheric CO_2 growth rate (dC/dt) is calculated from annual differences in mean concentration from an array of selected marine boundary layer sites. To estimate the uncertainty in dC/dt due to site selection, we construct 100 bootstrap simulations of alternative observing networks. The pre-1980 value of dC/dt is calculated as the annual difference in mean concentration at Mauna Loa and the South Pole, with an error structure derived from comparison with the contemporary marine boundary layer network extended back to 1959 (Methods).

For F_{Ff}, we use emission estimates from three global inventories—the Carbon Dioxide Information Analysis Center, BP and the Emissions Database for Global Atmospheric Research—to incorporate possible biases that may persist through the entire record, such as energy-to-carbon conversion factors. Each emission inventory is divided into two groups, developed nations (members of the Organization for Economic Co-operation and Development) and developing nations (non-members). The 2σ error for F_{Ff} is then estimated as 5% of emissions for all developed nations and 10% of emissions for all developing nations[20].

For F_{Li}, we use three independent inventories derived from model simulations of land-use and forest statistics[17,18,19], and assign a 2σ error of 50% to each F_{Li} inventory. Because F_{Ff} and F_{Li} inventory errors do not vary randomly from year to year, we simulate the uncertainty by generating 500 time series of autocorrelated errors (persistence of ~20 yr) for each inventory. For the case with only F_{Ff} we combine the three F_{Ff} inventories, and for the case with both F_{Ff} and F_{Li} we combine each of our F_{Ff} inventories with each of our F_{Li} inventories into a 3×3 matrix of emission scenarios, for a total of 4,500 ΣF emission simulations ($N = 3 \times 3 \times 500 = 4{,}500$). Values of AF and ΣN were calculated by combining each simulation of ΣF with a randomly chosen simulation of dC/dt. Unless otherwise noted, all uncertainty ranges correspond to 2σ.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 6 October 2011; accepted 6 June 2012.

1. Meehl, G. A. *et al.* in *Climate Change 2007: The Physical Science Basis* (eds Solomon, S. *et al.*) 792–802 (Cambridge Univ. Press, 2007).
2. Schimel, D. S. *et al.* Recent patterns and mechanisms of carbon exchange by terrestrial ecosystems. *Nature* **414,** 169–172 (2001).
3. Friedlingstein, P. *et al.* Climate-carbon cycle feedback analysis: results from the C4MIP model intercomparison. *J. Clim.* **19,** 3337–3353 (2006).
4. Pan, Y. *et al.* A large and persistent carbon sink in the world's forests. *Science* **333,** 988–993 (2011).
5. Piao, S. *et al.* Net carbon dioxide losses of northern ecosystems in response to autumn warming. *Nature* **451,** 49–52 (2008).
6. Zhao, M. & Running, S. W. Drought-induced reduction in global terrestrial net primary production from 2000 through 2009. *Science* **329,** 940–943 (2010).
7. McKinley, G. A., Fay, A. R., Takahashi, T. & Metzl, N. Convergence of atmospheric and North Atlantic carbon dioxide trends on multidecadal timescales. *Nature Geosci.* **4,** 606–610 (2011).
8. Le Quéré, C. *et al.* Saturation of the Southern Ocean CO_2 sink due to recent climate change. *Science* **316,** 1735–1738 (2007).
9. Schuster, U. & Watson, A. J. A variable and decreasing sink for atmospheric CO_2 in the North Atlantic. *J. Geophys. Res.* **112,** C1006 (2007).
10. Le Quéré, C., Takahashi, T., Buitenhuis, E. T., Rödenbeck, C. & Sutherland, S. C. Impact of climate change and variability on the global oceanic sink of CO_2. *Glob. Biogeochem. Cycles* **24,** GB4007 (2010).
11. Sarmiento, J. L. *et al.* Trends and regional distributions of land and ocean carbon sinks. *Biogeosciences* **7,** 2351–2367 (2010).
12. Knorr, W. Is the airborne fraction of anthropogenic CO_2 emissions increasing? *Geophys. Res. Lett.* **36,** L21710 (2009).
13. Gloor, M., Sarmiento, J. L. & Gruber, N. What can be learned about carbon cycle climate feedbacks from the CO_2 airborne fraction? *Atmos. Chem. Phys.* **10,** 7739–7751 (2010).
14. Boden, T. A., Marland, G. & Andres, R. J. Global, regional, and national fossil-fuel CO_2 emissions. *Carbon Dioxide Information Analysis Center* http://cdiac.ornl.gov/trends/emis/overview.html (2010).
15. BP. *Statistical Review of World Energy* http://www.bp.com/sectionbodycopy.do?categoryId=7500&contentId=7068481 2011).
16. European Commission, Emissions Database for Global Atmospheric Research (EDGAR). *Europa - EDGAR Overview* http://edgar.jrc.ec.europa.eu/overview.php?v=40 (2009).
17. Friedlingstein, P. *et al.* Update on CO_2 emissions. *Nature Geosci.* **3,** 811–812 (2010).
18. Stocker, B., Strassmann, K. & Joos, F. Sensitivity of Holocene atmospheric CO_2. *Biogeosciences* **7,** 921–952 (2011).
19. Yang, X., Richardson, T. K. & Jain, A. K. Contributions of secondary forest and nitrogen dynamics to terrestrial carbon uptake. *Biogeosciences* **7,** 3041–3050 (2010).
20. Marland, G., Hamal, K. & Jonas, M. How uncertain are estimates of CO_2 emissions? *J. Ind. Ecol.* **13,** 4–7 (2009).
21. Canadell, J. G. *et al.* Contributions to accelerating atmospheric CO_2 growth from economic activity, carbon intensity, and efficiency of natural sinks. *Proc. Natl Acad. Sci. USA* **104,** 18866–18870 (2007).
22. Le Quéré, C. *et al.* Trends in the sources and sinks of carbon dioxide. *Nature Geosci.* **2,** 831–836 (2009).

Acknowledgements A.P.B. was supported by the US National Research Council and the US National Science Foundation. This manuscript benefitted from comments from J. Neff, N. Lovenduski and G. Marland. We also thank K. Masarie for performing the bootstrap calculations on the atmospheric CO_2 sampling network. This work would not have been possible without the careful measurements made by scientists at NOAA ESRL and volunteer sample collectors throughout the world.

Author Contributions All authors identified the need for this analysis. P.P.T. and J.B.M. contributed to the uncertainty analysis, and P.P.T. and A.P.B. devised the Monte Carlo simulations. A.P.B. and C.B.A. wrote the paper with assistance from all other co-authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to A.P.B. (apballantyne@gmail.com).

LETTER

doi:10.1038/nature11300

Persistent near-tropical warmth on the Antarctic continent during the early Eocene epoch

Jörg Pross[1,2], Lineth Contreras[1], Peter K. Bijl[3], David R. Greenwood[4], Steven M. Bohaty[5], Stefan Schouten[6], James A. Bendle[7], Ursula Röhl[8], Lisa Tauxe[9], J. Ian Raine[10], Claire E. Huck[11], Tina van de Flierdt[11], Stewart S. R. Jamieson[12], Catherine E. Stickley[13], Bas van de Schootbrugge[1], Carlota Escutia[14], Henk Brinkhuis[3] & Integrated Ocean Drilling Program Expedition 318 Scientists*

The warmest global climates of the past 65 million years occurred during the early Eocene epoch (about 55 to 48 million years ago), when the Equator-to-pole temperature gradients were much smaller than today[1,2] and atmospheric carbon dioxide levels were in excess of one thousand parts per million by volume[3,4]. Recently the early Eocene has received considerable interest because it may provide insight into the response of Earth's climate and biosphere to the high atmospheric carbon dioxide levels that are expected in the near future[5] as a consequence of unabated anthropogenic carbon emissions[4,6]. Climatic conditions of the early Eocene 'greenhouse world', however, are poorly constrained in critical regions, particularly Antarctica. Here we present a well-dated record of early Eocene climate on Antarctica from an ocean sediment core recovered off the Wilkes Land coast of East Antarctica. The information from biotic climate proxies (pollen and spores) and independent organic geochemical climate proxies (indices based on branched tetraether lipids) yields quantitative, seasonal temperature reconstructions for the early Eocene greenhouse world on Antarctica. We show that the climate in lowland settings along the Wilkes Land coast (at a palaeolatitude of about 70° south) supported the growth of highly diverse, near-tropical forests characterized by mesothermal to megathermal floral elements including palms and Bombacoideae. Notably, winters were extremely mild (warmer than 10 °C) and essentially frost-free despite polar darkness, which provides a critical new constraint for the validation of climate models and for understanding the response of high-latitude terrestrial ecosystems to increased carbon dioxide forcing.

The climate and ecosystem evolution on Antarctica before the onset of continental-scale glaciation at the Eocene/Oligocene transition (~33.9 Myr ago) is still poorly resolved owing to the obliteration or coverage of potential archives by the Antarctic ice sheet. Available data are primarily based on records from the Antarctic Peninsula, which are only partly representative of climate and ecosystem conditions on the Antarctic mainland[7]. Terrestrial proxy data generally indicate cool temperate conditions supporting a vegetation dominated by podocarpaceous conifers during the Palaeocene epoch (~65–56 Myr ago) and southern beech (*Nothofagus*) during the middle Eocene epoch (~49–37 Myr ago), followed by the final demise of angiosperm-dominated woodlands as a result of Cenozoic cooling and the development of the Antarctic cryosphere around Eocene/Oligocene boundary times[8–10]. This virtually makes the terrestrial realm of the high southern latitudes a climatic *terra incognita* for the peak warming phase of the Cenozoic greenhouse world.

We apply terrestrial palynology and palaeothermometry based on the methylation index of branched tetraethers (MBT) and the cyclization ratio of branched tetraethers (CBT) to a new sedimentary record from the Wilkes Land margin, East Antarctica, recovered by the Integrated Ocean Drilling Program (IODP Expedition 318 Site U1356; see ref. 11 and Fig. 1). These data sets provide the framework for a terrestrial climate reconstruction for the early Eocene of Antarctica. The record presented here comprises a succession of mid-shelfal sediments with excellent chronostratigraphic control (Supplementary Fig. 1), representing early Eocene (53.6–51.9 Myr ago) greenhouse conditions and, separated by a ~2 Myr hiatus, an

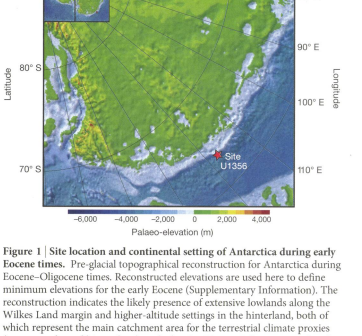


Figure 1 | Site location and continental setting of Antarctica during early Eocene times. Pre-glacial topographical reconstruction for Antarctica during Eocene–Oligocene times. Reconstructed elevations are used here to define minimum elevations for the early Eocene (Supplementary Information). The reconstruction indicates the likely presence of extensive lowlands along the Wilkes Land margin and higher-altitude settings in the hinterland, both of which represent the main catchment area for the terrestrial climate proxies (sporomorphs and biomarkers) studied at Site U1356. Palaeotopography after ref. 29; early Eocene coordinates obtained from the Ocean Drilling Stratigraphic Network after ref. 30.

[1]Palaeoenvironmental Dynamics Group, Institute of Geosciences, Goethe University Frankfurt, Altenhöferallee 1, 60438 Frankfurt, Germany. [2]Biodiversity and Climate Research Centre, Senckenberganlage 25, 60325 Frankfurt, Germany. [3]Biomarine Sciences, Institute of Environmental Biology, Laboratory of Palaeobotany and Palynology, Utrecht University, Budapestlaan 4, 3584 CD Utrecht, The Netherlands. [4]Biology Department, Brandon University, 270 18th Street, Brandon, Manitoba R7A 6A9, Canada. [5]Ocean and Earth Science, National Oceanography Centre Southampton, University of Southampton, European Way, Southampton SO14 3ZH, UK. [6]Department of Marine Organic Biogeochemistry, NIOZ Royal Netherlands Institute of Sea Research, Post Office 59, 1790 AB Den Burg (Texel), The Netherlands. [7]Glasgow Molecular Organic Geochemistry Laboratory, School of Geographical and Earth Sciences, University of Glasgow, Glasgow G12 8QQ, UK. [8]MARUM – Center for Marine Environmental Sciences, University of Bremen, Leobener Straße, 28359 Bremen, Germany. [9]Scripps Institution of Oceanography, University of California, San Diego, La Jolla, California 92093, USA. [10]Department of Palaeontology, GNS Science, PO Box 30368, Lower Hutt 6009, New Zealand. [11]Imperial College London, Department of Earth Science and Engineering, South Kensington Campus, Exhibition Road, London SW7 2AZ, UK. [12]Department of Geography, Durham University, Science Laboratories, South Road, Durham DH1 3LE, UK. [13]Department of Geology, University of Tromsø, 9037 Tromsø, Norway. [14]Instituto Andaluz de Ciencias de la Tierra, Avenida de las Palmeras, 4 18100 Armilla (Granada), Spain.

*Lists of participants and affiliations appear at the end of the paper.

interval of cooling presumed within the latest early Eocene to middle Eocene (49.3–46 Myr ago; here informally referred to as the 'mid-Eocene'). Palynological and geochemical evidence independently supports the contention that the Wilkes Land sector of Antarctica is indeed the source region for the Eocene terrestrial palynomorphs and biomarkers present in the sediment core from Site U1356 (Supplementary Information).

Non-metric multidimensional scaling techniques show that the Eocene sporomorph assemblages at Site U1356 represent two main biomes (Fig. 2 and Supplementary Information). A highly diverse paratropical rainforest biome prevailed during the early Eocene, probably occupying the coastal lowlands of the Wilkes Land margin. This biome includes numerous mesothermal to megathermal taxa characteristic of modern subtropical to tropical settings in Australia, New Guinea and New Caledonia[12]. In addition to ferns and tree ferns (*Lygodium*, Cyatheaceae), it is characterized by the presence of palms (Arecaceae), Bombacoideae (Malvaceae), *Strasburgeria* (Strasburgeriaceae), *Beauprea* (Proteaceae), *Anacolosa* (Olacaceae) and *Spathiphyllum* (Araceae) (Fig. 2). Although these additional taxa occur only in low abundance, their presence is highly significant. Because they are pollinated by insects, their pollen dispersal in extant rainforests is generally restricted to less than 100 m (ref. 13). Hence, even low percentages of their pollen in the Site U1356 record indicate that these plants formed a substantial part of the Wilkes Land margin vegetation.

The palm and Bombacoideae pollen not only represent the southernmost documented occurrences for both taxa during the Eocene, but, importantly, imply that winter temperatures remained substantially above freezing. Extant palms occur naturally only in regions with a coldest-month mean temperature (CMMT) of ≥5 °C (ref. 1). Because their cold-season temperature requirements increase further when palms grow under a high partial pressure of atmospheric CO_2, the CMMT implied by palms during the early Eocene greenhouse world was at least 8 °C (ref. 14). Even warmer conditions are suggested by the record of Bombacoideae, which today occur where CMMT > 10 °C. Because even the most winter-hardy extant palms are severely damaged by short-term freezing, with a series of consecutive years of unfavourable climate eventually being lethal[15], winters must have been essentially frost-free.

Sporomorphs representing a lower-diversity temperate rainforest biome, with taxa characteristic of extant forests in montane settings of Australia, New Caledonia, New Guinea and New Zealand[12], typically account for ~30% of sporomorphs during the early Eocene. Characteristic taxa include *Nothofagus* (*fusca* type), Araucariaceae, Proteaceae and *Podocarpus*; mesothermal to megathermal, frost-sensitive taxa are consistently absent. Judging from its floral composition, this temperate rainforest biome occupied cooler environments of Wilkes Land located farther inland and/or at higher elevations, and therefore provides insight into the climate conditions deeper within the Antarctic continent. The coeval existence of a temperate rainforest biome in the hinterland and a paratropical rainforest in the lowlands of the Wilkes Land margin indicates a pronounced continental interior-to-coastal temperature gradient during the early Eocene.

A markedly different vegetation pattern is documented for the mid-Eocene time interval, with a strong expansion of the *Nothofagus*-dominated temperate rainforest biome and the near-extirpation of the paratropical rainforest biome; notably, the remainder of the latter biome is devoid of megathermal elements (Fig. 2). Hence, our data suggest that the temperate rainforest biome became dominant over the entire catchment area of Site U1356, also extending into the coastal regions, and that relict mesothermal components of the paratropical rainforest biome persisted only in localized pockets along the Wilkes Land margin. These shifts in dominance and floral composition indicate a strong cooling, which in light of the cold-season sensitivity of meso- and megathermal taxa was particularly pronounced in winter temperatures, and a strong weakening of the temperature gradient between coastal and montane regions of the Wilkes Land margin.

To quantify further the sporomorph-derived palaeoclimatic information, we carried out bioclimatic analyses using the nearest living relative concept[16] to reconstruct the mean annual temperature (MAT), the mean winter and summer temperatures (MWT and MST)

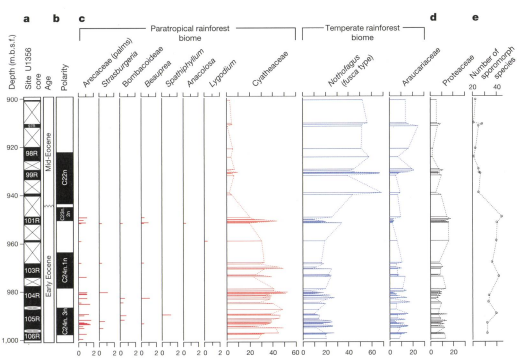


Figure 2 | Data from Site U1356 for the early Eocene to mid-Eocene. a, Core recovery. m.b.s.f., metres below sea floor. **b**, Geological age[11]. **c**, Relative abundances of selected sporomorphs representative of the paratropical and temperate rainforest biomes. **d**, Relative abundances of Proteaceae pollen. Data based on samples with counts of ≥90 specimens. **e**, Number of sporomorph species rarefied at 280 grains. The number of sporomorph species from the early Eocene is significantly higher than that from the mid-Eocene (Mann–Whitney test, $P < 0.000005$).

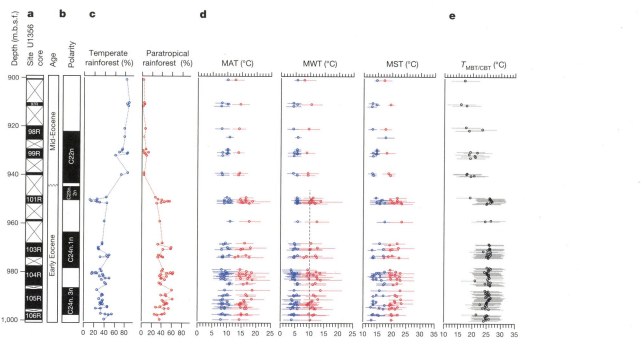


Figure 3 | Climate reconstruction for the Wilkes Land sector of Antarctica during the early and mid-Eocene derived from Site U1356. a, Core recovery. **b**, Geological age[11]. **c**, Relative abundances of sporomorphs representing the temperate and paratropical rainforest biomes. **d**, Estimates of MAT, MWT and MST for the temperate (blue) and paratropical (red) rainforest biomes, based on the methodology of ref. 16. Error bars represent the minimum and maximum estimates determined using that method. The vertical dashed line marks the minimum requirements of Bombacoideae for the mean temperature of the coldest month. **e**, Temperatures derived from the MBT/CBT index, with horizontal error bars indicating the calibration standard error (±5 °C). This error refers to absolute temperature estimates across all environmental settings of the modern calibration; thus, the error of the within-record variation is much smaller. Relative sporomorph abundances and sporomorph-based climate estimates are based on samples with counts of ≥90 specimens.

(Fig. 3), and the mean annual precipitation (Supplementary Fig. 5). These results were critically assessed through a comparison with reconstructions using a different methodology that also relies on the nearest living relative concept (the coexistence approach of ref. 17; see Supplementary Information). Because the two recognized biomes represent distinct environments with different climatic conditions, our approach allows a spatiotemporally differentiated view of the climate evolution of Wilkes Land from early Eocene peak warmth through the onset of mid-Eocene cooling. Our temperature estimates for the paratropical rainforest biome show that climate along the Wilkes Land margin was generally warm until at least 51.9 Myr ago. Most samples indicate temperatures of 16 ± 5 °C for MAT, 11 ± 5 °C for MWT and 21 ± 5 °C for MST, although a small number also yield colder or warmer values (Fig. 3). A markedly cooler climate emerges for the temperate rainforest biome, in particular for MAT and MWT, for which most samples yield values of 9 ± 3 °C and 5 ± 2 °C, respectively. For MST, the data show a strong scatter between 14 ± 1 °C and 18 ± 3 °C, and the values overlap partly with those for the paratropical rainforest biome. For both biomes, the mean annual precipitation was persistently more than 100 cm yr^{-1} (Supplementary Information).

For the mid-Eocene interval, our reconstructions based on the relicts of the paratropical rainforest biome suggest a pronounced cooling, although this trend is partly within the error limits of the data. The estimated MAT is 14 ± 3 °C, which represents a decline of ~2 °C from the early Eocene. Our data also indicate a decline in MWT and MST, although these trends are again within the error limits. Temperatures reconstructed for the temperate rainforest biome are comparable to those from the early Eocene, which is consistent with there being no major changes in the composition of this biome between both intervals.

Independent support for a warm terrestrial climate during the early Eocene and marked cooling during the mid-Eocene comes from our MBT/CBT palaeothermometry data (Fig. 3). Soil temperatures of ~24–27 °C are estimated for the early Eocene, and ~17–20 °C for the mid-Eocene. These temperatures fall close to the MSTs derived for the paratropical rainforest biome. This suggests that the branched tetraethers in Site U1356 sediments originated from coastal lowland soils of the Wilkes Land sector of Antarctica and could imply a bias of the MBT/CBT proxy towards summer temperatures, although such a bias has not been observed in modern mid-latitude climates (Supplementary Information).

Our data, which provide continental temperature reconstructions for the high southern latitudes during the early Eocene greenhouse world, show that paratropical conditions persisted in the lowlands of the Wilkes Land margin of Antarctica from at least 53.9 to 51.9 Myr ago. Notably, our estimates yield a constraint on Antarctic winter temperatures during peak greenhouse conditions. The CMMT and MWT estimates of ≥10 °C and 11 ± 5 °C, respectively, compare favourably with deep-water temperatures of ~11 °C in the marine realm at this time[6,18]. Because early Eocene deep waters were sourced from downwelling surface waters in the high southern latitudes off Antarctica[19], winter temperatures in these regions cannot have dropped much below 11 °C. Although our MWT estimates are not representative of the Antarctic continent as a whole, they bear implications for the current debates on the general ability of climate models to reproduce extreme greenhouse conditions and the response of polar ecosystems to increased CO_2 forcing.

When run with conservative estimates of atmospheric CO_2 levels for the early Eocene, fully coupled climate models yield high-latitude terrestrial winter temperatures considerably below freezing[20], and they produce warm (that is, above-freezing) winters in the terrestrial high latitudes only when radiative forcing is strongly enhanced[21]. Hence, our winter temperatures for Wilkes Land provide a critical reference point for understanding the climate dynamics of the early Eocene greenhouse world. They are in remarkably close agreement with simulated MWTs for the Wilkes Land region when radiative forcings equivalent to 2,240 p.p.m.v. and 4,480 p.p.m.v. CO_2 are applied (ref. 21), suggesting that enhancing radiative forcing in models may help resolve the persistent data–model mismatch. However, factors other than extremely high atmospheric greenhouse gas forcing may have contributed to the winter warmth along the Wilkes Land sector of Antarctica. They

include winter cloud radiative forcing over high-latitude land masses[22], possibly connected to high ocean-to-land moisture transport[23]. Our precipitation estimates (Supplementary Fig. 5) and the presence of rainforest biomes consistently suggest high moisture availability throughout the year, thus lending support for this mechanism being in operation in the Wilkes Land sector of Antarctica. A high moisture flux from the ocean was facilitated by the presence of extremely warm surface waters in the Australo-Antarctic gulf, resulting from the subtropically derived, clockwise-flowing proto-Leeuwin current[24]. Warm surface waters off Wilkes Land are documented by mass occurrences of the subtropical dinoflagellate cyst *Apectodinium*[11].

Our data also provide new insights into the physiological ecology of high-latitude forests, which are subject to seasonally extreme changes in light levels. The ~50 days of polar darkness on the Wilkes Land margin poses severe constraints on the plants' carbon gain by photosynthesis and carbon loss by respiration. Because carbon loss by respiration typically increases with temperature[25], it has been argued that polar winters must have been cool rather than warm[26]. Our MBT/CBT temperature data, which under the most conservative (that is, 'coldest') assumption represent MST, are typically between 24 and 27 °C for the early Eocene (Fig. 3). They are similar to, although possibly slightly warmer than, the terrestrial MST predicted by climate models using high radiative forcing (20–25 °C; ref. 21). Over a wide range of CO_2 forcing, the models yield a temperature seasonality of the order of 10 °C, thus suggesting a MWT of 10–15 °C. This evidence, which is strictly independent of our vegetation-based climate reconstructions, contradicts the scenario of cold winters on Wilkes Land, therefore suggesting that respiration losses under a highly seasonal polar light regime were compensated for by a factor other than temperature. We suggest that the high atmospheric CO_2 levels of the early Eocene greenhouse climate were a decisive factor in the physiological ecology of high-latitude forests, most probably through causing a reduction in carbon respiration during the polar winter[27] and an increase in photosynthetic carbon gain during the growing season[28].

Our new data from the peak early Eocene greenhouse world indicate that a highly diverse forest vegetation containing evergreen elements can successfully colonize high-latitude, warm winter environments when atmospheric CO_2 levels are high. Depending on the thresholds in atmospheric CO_2 required by such plants, the duration of polar winters and the temperatures at which such forcing factors become significant, these results have important implications for the composition of high-latitude terrestrial ecosystems in a future anthropogenic greenhouse world with high atmospheric CO_2 levels and drastic polar amplification of warming.

METHODS SUMMARY

Palynology. Between 10 and 15 g of sediment was processed per sample. The dried sediment was weighed and spiked with *Lycopodium* spores to facilitate the calculation of absolute palynomorph abundances. Chemical processing comprised treatment with 30% HCl and 38% HF for carbonate and silica removal, respectively. Ultrasonication was used to disintegrate palynodebris. Residues were sieved over a 10-μm mesh and mounted on microscope slides, which were analysed at ×200 and ×1,000 magnification. A detailed, step-by-step processing protocol is given in Supplementary Information.

Sporomorph-based climate reconstructions. Bioclimatic analyses were carried out following ref. 16, but with data sources including Southern Hemisphere taxa, allowing the development of climatic profiles for each taxon as described in Supplementary Information. The results of the bioclimatic analyses were critically assessed through the application of the coexistence approach[17] to the data set using the same underlying database. Supplementary Table 1 lists all taxa that were evaluated through the bioclimatic analyses and the coexistence approach, their botanical affinity and the nearest living relatives used in the analyses.

Organic geochemistry. For MBT/CBT analyses, freeze-dried, powdered samples were extracted with an accelerated solvent extractor using a 9:1 (v/v) dichloromethane (DCM):methanol solvent mixture. The obtained extracts were separated over an activated Al_2O_3 column, using 9:1 (v/v) hexane:DCM, 1:1 (v/v) hexane: DCM, 1:1 (v/v) ethylacetate:DCM and 1:1 (v/v) DCM:methanol, into apolar, ketone, ethylacetate and polar fractions, respectively. The polar fractions containing the branched tetraether lipids were analysed by HPLC/APCI-MS (high-performance liquid chromatography/atmospheric pressure chemical ionization mass spectrometry) using an Agilent 1100 LC/MSD SL. MBT/CBT indices were calculated and converted into temperature estimates as described in Supplementary Information.

Received 28 February; accepted 8 June 2012.

1. Greenwood, D. R. & Wing, S. L. Eocene continental climates and latitudinal temperature gradients. *Geology* **23,** 1044–1048 (1995).
2. Bijl, P. K. *et al.* Early Palaeogene temperature evolution of the southwest Pacific Ocean. *Nature* **461,** 776–779 (2009).
3. Yapp, C. J. $Fe(CO_3)OH$ in goethite from a mid-latitude North American oxisol: estimate of atmospheric CO_2 concentration in the Early Eocene "climatic optimum". *Geochim. Cosmochim. Acta* **68,** 935–947 (2004).
4. Beerling, D. J. & Royer, D. L. Convergent Cenozoic CO_2 history. *Nat. Geosci.* **4,** 418–420 (2011).
5. Meinshausen, M. *et al.* The RCP greenhouse gas concentrations and their extensions from 1765 to 2300. *Clim. Change* **109,** 213–241 (2011).
6. Zachos, J. C., Dickens, G. R. & Zeebe, R. E. An early Cenozoic perspective on greenhouse warming and carbon-cycle dynamics. *Nature* **451,** 279–283 (2008).
7. Anderson, J. B. *et al.* Progressive Cenozoic cooling and the demise of Antarctica's last refugium. *Proc. Natl Acad. Sci. USA* **108,** 11356–11360 (2011).
8. Francis, J. E. Antarctic palaeobotany: clues to climate change. *Terra Antartica* **3,** 135–140 (1996).
9. Poole, I., Cantrill, D. & Utescher, T. A multi-proxy approach to determine Antarctic terrestrial palaeoclimate during the Late Cretaceous and Early Tertiary. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **222,** 95–121 (2005).
10. Truswell, E. M. & Macphail, M. K. Polar forests on the edge of extinction: what does the fossil spore and pollen evidence from East Antarctica say? *Aust. Syst. Bot.* **22,** 57–106 (2009).
11. Expedition 318 Scientists in *Proc. IODP*, Vol. 318 (eds Escutia, C., Brinkhuis, H., Klaus, A., and the Expedition 318 Scientists) (Integrated Ocean Drilling Program Management International, Inc., 2011).
12. Kershaw, A. P. in *Vegetation History* (eds Huntley, B. & Webb, T.) 237–306 (Kluwer, 1988).
13. Bush, M. B. & Rivera, R. Pollen dispersal and representation in a neotropical rain forest. *Glob. Ecol. Biogeogr.* **7,** 379–392 (1998).
14. Royer, D. L., Osborne, C. P. & Beerling, D. J. High CO_2 increases the freezing sensitivity of plants: implications for paleoclimatic reconstructions from fossil floras. *Geology* **30,** 963–966 (2002).
15. Larcher, W. & Winter, A. Frost susceptibility of palms: experimental data and their interpretation. *Principes* **25,** 143–152 (1981).
16. Greenwood, D. R., Archibald, S. B., Mathewes, R. W. & Moss, P. T. Fossil biotas from the Okanagan Highlands, southern British Columbia and northeastern Washington State: climates and ecosystems across an Eocene landscape. *Can. J. Earth Sci.* **42,** 167–185 (2005).
17. Mosbrugger, V. & Utescher, T. The coexistence approach – a method for quantitative reconstructions of Tertiary terrestrial palaeoclimate data using plant fossils. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **134,** 61–86 (1997).
18. Lear, C. H., Elderfield, H. & Wilson, P. A. Cenozoic deep-sea temperatures and global ice volumes from Mg/Ca in benthic foraminiferal calcite. *Science* **287,** 269–272 (2000).
19. Thomas, D. J., Bralower, T. J. & Jones, C. E. Neodymium isotopic reconstruction of late Paleocene-early Eocene thermohaline circulation. *Earth Planet. Sci. Lett.* **209,** 309–322 (2003).
20. Shellito, C. J., Sloan, L. C. & Huber, M. Climate model sensitivity to atmospheric CO_2 levels in the Early-Middle Paleogene. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **193,** 113–123 (2003).
21. Huber, M. & Caballero, R. The early Eocene equable climate problem revisited. *Clim. Past* **7,** 603–633 (2011).
22. Abbot, D. S., Huber, M., Bousquet, G. & Walker, C. C. High-CO_2 cloud radiative forcing feedback over both land and ocean in a global climate model. *Geophys. Res. Lett.* **36** (2009).
23. Abbot, D. S. & Tziperman, E. Sea ice, high-latitude convection, and equable climates. *Geophys. Res. Lett.* **35** (2008).
24. Huber, M. *et al.* Eocene circulation of the Southern Ocean: was Antarctica kept warm by subtropical waters? *Paleoceanography* **19** (2004).
25. Tjoelker, M. G., Oleksyn, J. & Reich, P. B. Modelling respiration of vegetation: evidence for a general temperature-dependent Q10. *Glob. Change Biol.* **7,** 223–230 (2001).
26. Read, J. & Francis, J. Responses of some Southern Hemisphere tree species to a prolonged dark period and their implications for high-latitude Cretaceous and Tertiary floras. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **99,** 271–290 (1992).
27. Beerling, D. J. & Osborne, C. P. Physiological ecology of Mesozoic polar forests in a high CO_2 environment. *Ann. Bot. (Lond.)* **89,** 329–339 (2002).
28. DeLucia, E. H. *et al.* Net primary production of a forest ecosystem with experimental CO_2 enrichment. *Science* **284,** 1177–1179 (1999).
29. Wilson, D. S. *et al.* Antarctic topography at the Eocene-Oligocene boundary. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **335–336,** 24–34 (2011).
30. Hay, W. W. *et al.* Alternative global Cretaceous paleogeography. *Spec. Publ. Geol. Soc. Am.* **332,** 1–47 (1999).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This research used samples and data provided by the Integrated Ocean Drilling Program (IODP). The IODP is sponsored by the US National Science Foundation and participating countries under the management of Joint Oceanographic Institutions, Inc. Financial support for this research was provided by the German Research Foundation, to J.P. (grant PR 651/10) and U.R. (grant RO 1113/6); the Biodiversity and Climate Research Center of the Hessian Initiative for Scientific and Economic Excellence, to J.P.; the Netherlands Organisation for Scientific Research, to H.B. and S.S. (VICI grant); the Natural Sciences and Engineering Research Council of Canada, to D.R.G. (grant DG 311934); the Natural Environment Research Council, to S.M.B. (grant Ne/J019801/1). J.A.B. (grant Ne/I00646X/1) and T.v.d.F. (grant Ne/I006257/1); the US National Science Foundation, to L.T. (grant OCE 1058858); and the New Zealand Ministry of Science and Innovation, to J.I.R.. We thank J. Francis, S. Gollner and M. Huber for discussions, and B. Coles, E. Hopmans, K. Kreissig, A. Mets, J. Ossebaar, B. Schminke, J. Treehorn and N. Welters for technical support.

Author Contributions J.P., H.B. and S.S. designed the research. L.C., J.P. and J.I.R. analysed the sporomorphs, D.R.G. and L.C. conducted the quantitative climate reconstructions, and P.K.B. and S.S. carried out the MBT/CBT analyses. All authors contributed to the interpretation of the data. J.P. wrote the paper with input from all authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to J.P. (joerg.pross@em.uni-frankfurt.de).

Integrated Ocean Drilling Program Expedition 318 Scientists Henk Brinkhuis[1], Carlota Escutia Dotti[2], Adam Klaus[3], Annick Fehr[4], Trevor Williams[5], James A. P. Bendle[6], Peter K. Bijl[1], Steven M. Bohaty[7], Stephanie A. Carr[8], Robert B. Dunbar[9], Jhon J. González[2], Travis G. Hayden[10], Masao Iwai[11], Francisco J. Jimenez-Espejo[12†], Kota Katsuki[13], Gee Soo Kong[14], Robert M. McKay[15], Mutsumi Nakai[16], Matthew P. Olney[17], Sandra Passchier[18], Stephen F. Pekar[19], Jörg Pross[20,21], Christina R. Riesselman[22], Ursula Röhl[23], Toyosaburo Sakai[24], Prakash K. Shrivastava[14], Catherine E. Stickley[25], Saiko Sugisaki[26†], Lisa Tauxe[27], Shouting Tuo[28], Tina van de Flierdt[29], Kevin Welsh[3d] & Masako Yamane[11]

Affiliations for participants: [1]Biomarine Sciences, Institute of Environmental Biology, Laboratory of Palaeobotany and Palynology, Utrecht University, Budapestlaan 4, 3584 CD Utrecht, The Netherlands. [2]Instituto Andaluz de Ciencias de la Tierra, CSIC-Universite de Granada, Campus de Fuentenueva s/n, 18002 Granada, Spain. [3]United States Implementing Organization, Integrated Ocean Drilling Program, Texas A&M University, 1000 Discovery Drive, College Station, Texas 77845, USA. [4]Aachen University, Institute for Applied Geophysics and Geothermal Energy, Mathieustraße 6, D-52074 Aachen, Germany. [5]Borehole Research Group, Lamont-Doherty Earth Observatory of Columbia University, PO Box 1000, 61 Route 9W, Palisades, New York 10964, USA. [6]Geographical and Earth Sciences, University of Glasgow, G128QQ Glasgow, UK. [7]Ocean and Earth Science, National Oceanography Centre Southampton, University of Southampton, European Way, SO14 3ZH Southampton, UK. [8]Department of Chemistry and Geochemistry, Colorado School of Mines, 1500 Illinois Street, Golden, Colorado 80401, USA. [9]Department of Environmental Earth System Science, Stanford University, 325 Braun Hall, Building 320, Stanford, California 94305-2115, USA. [10]Department of Geology, Western Michigan University, 1187 Rood Hall, 1903 West Michigan Avenue, Kalamazoo, Michigan 49008, USA. [11]Department of Natural Science, Kochi University, 2-5-1 Akebono-cho, Kochi 780-8520, Japan. [12]Institute for Research on Earth Evolution, Japan Agency for Marine-Earth Science and Technology, Natsushima-cho 2-15, Yokosuka 237-0061, Japan. [13]Marine Center for Advanced Core Research, Kochi University, B200 Monobe, Nankoku, Kochi 783-8502, Japan. [14]Petroleum and Marine Research Division, Korea Institute of Geoscience and Mineral Resources, 30 Gajeong-dong, Yuseong-gu, Daejeon 305-350, Republic of Korea. [15]Antarctic Research Centre, Victoria University of Wellington, PO Box 600, Wellington 6140, New Zealand. [16]Education Department, Daito Bunka University, 1-9-1 Takashima-daira, Itabashi-ku, Tokyo 175-8571, Japan. [17]Department of Geology, University of South Florida, Tampa, 4202 East Fowler Avenue, SCA 528, Tampa, Florida 33620, USA. [18]Earth and Environmental Studies, Montclair State University, 252 Mallory Hall, 1 Normal Avenue, Montclair, New Jersey 07043, USA. [19]School of Earth and Environmental Sciences, Queens College, 65-30 Kissena Boulevard, Flushing, New York 11367, USA. [20]Paleoenvironmental Dynamics Group, Institute of Geosciences, Goethe University Frankfurt, Altenhöferallee 1, 60438 Frankfurt, Germany. [21]Biodiversity and Climate Research Centre, Senckenberganlage 25, 60325 Frankfurt, Germany. [22]MARUM – Center for Marine Environmental Sciences, University of Bremen, Leobener Straße, 28359 Bremen, Germany. [23]Department of Geology, Utsunomiya University, 350 Mine-Machi, Utsunomiya 321-8505, Japan. [24]Antarctica Division, Geological Survey of India, NH5P, NIT, Faridabad 121001, Hariyana, India. [25]Department of Geology, Universitet i Tromsø, N-9037 Tromsø, Norway. [26]Department of Polar Science, Graduate University of Advanced Study, 10-3 Midori-cho, Tachikawa City, Tokyo 190-8518, Japan. [27]Scripps Institution of Oceanography, University of California, San Diego, La Jolla, California 92093-0220, USA. [28]School of Ocean and Earth Science, Tongji University, 1239 Spring Road, Shanghai 200092, China. [29]Department of Earth Science and Engineering, Imperial College London, London SW7 2AZ, UK. [30]School of Earth Sciences, University of Queensland, St Lucia, Brisbane, Queensland 4072, Australia. [31]Earth and Planetary Science, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan. †Present addresses: Instituto Andaluz de Ciencias de la Tierra, CSIC-Universidad de Granada, Armilla, 18100 Granada, Spain (F.J.J.-E.); US Geological Survey, Eastern Geology and Paleoclimate Science Center, 926A National Center, Reston, Virginia 20192, USA (C.R.R.); Japan Agency for Marine-Earth Science and Technology, Frontier Building 4F, 2-15 Natsushima-cho, Yokosuka City, Kanagawa 237-0061, Japan (S.S.).

LETTER

doi:10.1038/nature11226

Universal species–area and endemics–area relationships at continental scales

David Storch[1,2], Petr Keil[3] & Walter Jetz[3]

Despite the broad conceptual and applied relevance of how the number of species or endemics changes with area (the species–area and endemics–area relationships (SAR and EAR)), our understanding of universality and pervasiveness of these patterns across taxa and regions has remained limited. The SAR has traditionally been approximated by a power law[1], but recent theories predict a triphasic SAR in logarithmic space, characterized by steeper increases in species richness at both small and large spatial scales[2–6]. Here we uncover such universally upward accelerating SARs for amphibians, birds and mammals across the world's major landmasses. Although apparently taxon-specific and continent-specific, all curves collapse into one universal function after the area is rescaled by using the mean range sizes of taxa within continents. In addition, all EARs approximately follow a power law with a slope close to 1, indicating that for most spatial scales there is roughly proportional species extinction with area loss. These patterns can be predicted by a simulation model based on the random placement of contiguous ranges within a domain. The universality of SARs and EARs after rescaling implies that both total and endemic species richness within an area, and also their rate of change with area, can be estimated by using only the knowledge of mean geographic range size in the region and mean species richness at one spatial scale.

The scale dependence of species richness has implications for all biodiversity patterns[1]. The SAR has been used to extrapolate species richness across spatial scales and also to estimate species extinctions after habitat loss[7,8] (but see refs 9, 10), typically relying on its particular universal properties. However, the universality of the shape of the SAR has been questioned[11]. The nested SARs (in which smaller sample areas are located within larger ones) are classically described as a power law across most spatial scales[1,12], but current theoretical approaches predict that species richness first increases steeply with area at a decelerating rate, then increases roughly linearly in logarithmic space, and accelerates upwards again when sample areas approach the size of individual species' geographic ranges[2–6]. In contrast to the well-documented curvature over small areas[13,14], data availability has so far hindered generalizations about the SAR at large scales. The EAR is the relationship between the area of a region and the number of species restricted (that is, endemic) to it. The EAR provides information on the number of species that may go extinct if parts of the area are destroyed or transformed[9,15,16], because being endemic to the area would imply that any local extinction is also global. Despite the potential of the EAR in biodiversity science and conservation[9,15,17], its empirical shape at biogeographic scales has remained largely undocumented. The slope of the EAR at smaller spatial scales is expected to be connected to the slope of the SAR at large scales, because an increase in species richness with increasing study plot area (the SAR) corresponds to a decrease in the number of species that are restricted (that is, endemic) to the remaining area[16] (that is, the area not included in the study plot; see Supplementary Discussion and Supplementary Figs 1 and 2).

Here we provide a construction of fully nested continental SARs and EARs for all amphibians, birds and mammals (see refs 18 and 19 for data description and validation, and Methods and Supplementary Table 1 for details). SARs for all continents and taxa accelerate upward in log–log space (Fig. 1a–c). Differences in the SAR position along the *y* axis correspond to known differences in total species richness of individual continents and taxa[20,21]; for example, birds have consistently

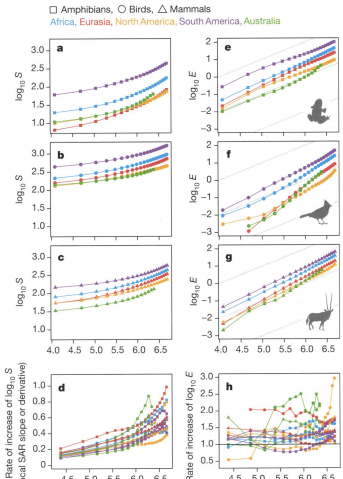


Figure 1 | SARs and EARs across five continents and three vertebrate classes. a–**c**, **e**–**g**, The SARs for amphibians (**a**), birds (**b**) and mammals (**c**) reveal an upward-accelerating shape for logarithmic axes, whereas EARs for amphibians (**e**), birds (**f**) and mammals (**g**) are more or less linear. **d**, **h**, Confirmation by plotting the local slopes (derivatives) of the relationships for each continent (**d** for SARs and **h** for EARs). All relationships were constructed by using a strictly nested quadrat design. Grey lines correspond to a power law with a slope of 1; that is, proportionality between area and the number of species. *S* is the mean number of species, *E* is the mean number of endemics, and *A* is the area in km^2.

[1]Center for Theoretical Study, Charles University and the Academy of Sciences of the Czech Republic, Jilská 1, 110 00, Praha 1, Czech Republic. [2]Department of Ecology, Faculty of Science, Charles University, Viničná 7, 128 44 Praha 2, Czech Republic. [3]Department of Ecology and Evolutionary Biology, Yale University, 165 Prospect Street, New Haven, Connecticut 06520-8106, USA.

Table 1 | Slopes of the EARs and SARs calculated by using the nested quadrat design

Taxon	Continent	EAR slope	SAR slope (lower half)	SAR slope (upper half)
Birds	Eurasia	1.92 (1.70–2.06)	0.21 (0.19–0.22)	0.41 (0.37–0.45)
	Africa	1.39 (1.22–1.65)	0.21 (0.19–0.23)	0.40 (0.34–0.45)
	N. America	1.25 (1.00–1.61)	0.19 (0.17–0.20)	0.29 (0.27–0.32)
	S. America	1.30 (1.15–1.44)	0.17 (0.16–0.19)	0.39 (0.35–0.43)
	Australia	1.96 (1.20–2.67)	0.15 (0.12–0.17)	n.a.
Mammals	Eurasia	1.36 (1.26–1.43)	0.26 (0.24–0.27)	0.48 (0.44–0.52)
	Africa	1.24 (1.15–1.39)	0.23 (0.21–0.25)	0.48 (0.42–0.54)
	N. America	1.26 (1.11–1.43)	0.21 (0.19–0.22)	0.40 (0.34–0.45)
	S. America	1.21 (1.10–1.36)	0.17 (0.16–0.18)	0.44 (0.41–0.46)
	Australia	1.46 (1.09–1.85)	0.21 (0.19–0.23)	n.a.
Amphibians	Eurasia	1.15 (1.05–1.31)	0.35 (0.31–0.40)	0.70 (0.54–0.86)
	Africa	1.13 (1.05–1.25)	0.29 (0.26–0.32)	0.64 (0.53–0.75)
	N. America	1.00 (0.84–1.16)	0.26 (0.22–0.31)	0.58 (0.43–0.70)
	S. America	0.93 (0.86–1.02)	0.26 (0.24–0.28)	0.58 (0.49–0.68)
	Australia	1.12 (0.80–1.53)	0.27 (0.23–0.33)	n.a.

The slopes were estimated by using linear regression on logarithms of the mean number of species for each area (also logarithmically transformed). The EAR slopes were calculated across the whole range of areas. The SAR slopes were calculated separately for the lower and upper half of the analysed areas (cut-off: $\log_{10}$area = 6.1) to provide measures of both the lower and upper ends of the upward-accelerating SARs. For local slope estimates at each area see Fig. 1d. To give a general representation of the possible range of slopes that would be detected if biodiversity data were incomplete, we randomly selected only 10% of all possible positions of sampling windows, repeated the procedure 500 times and estimated the lower and upper 95% quantiles of the slopes obtained from the resampled data (see Methods).

higher species richness for a given area than do mammals, whereas amphibians typically show low richness. However, amphibians also show much steeper SARs than other taxa (Fig. 1d), and Eurasia has the steepest SARs for all taxa. An assessment of local slopes (derivatives) of the SARs illustrates the upward-increasing nature of the logarithmic SAR (Fig. 1d and Table 1). Considerable differences in this increase appear among taxa, most clearly between Eurasian amphibians and North American birds. We do not find evidence for the first phase of the triphasic SARs, which confirms the expectation that this phase occurs only when the number of individuals becomes limited[4,11]; that is, at scales considerably finer than those made possible by the current grain size of global distribution data[19].

The nonlinear shapes of SARs stand in striking contrast to those observed for EARs (Fig. 1e–g). All continents and taxa show a consistent and seemingly linear increase in number of endemics with increasing area in logarithmic space. Local slopes of EARs are reasonably invariant with scale, taxon and continent (Fig. 1h), although some show a slight increase at areas above 3×10^6 km^2. Except for generally steeper slopes in birds, EAR slopes tend to vary between 0.75 and 1.5 and are often close to 1 (Fig. 1h and Table 1), indicating that the number of endemic species increases more or less proportionally with area.

The increasing slope of the SAR at large spatial scales is predicted to be associated with increasing species spatial turnover as sample areas approach the sizes of species' geographic ranges[2–4] (see Supplementary Discussion). We contend that the SAR curvature may thus be dependent on an 'effective' range size equal to the mean range size. Because of its similar foundation, we predict that EARs will show similar range size dependence. We therefore rescaled all area axes such that one areal unit corresponds to the mean species geographic range for a given continent and taxon:

$$A_r = A/\overline{R}_{t,c} \quad (1)$$

where A_r is the rescaled area, A is the area of the study plot and $\overline{R}_{t,c}$ is the mean range size for taxon t and continent c. In addition we rescaled the vertical axis to represent species richness proportional to the richness of an area equal to $\overline{R}_{t,c}$:

$$S_r = S_A/S_{\overline{R}(t,c)} \quad (2)$$

and

$$E_r = E_A/E_{\overline{R}(t,c)} \quad (3)$$

where S_r and E_r are the rescaled counts of species and endemics, S_A and E_A are mean counts for a given area, and $S_{\overline{R}(t,c)}$ and $E_{\overline{R}(t,c)}$ are mean richness values for the area that equals the mean geographic range size of a given taxon and continent. Under this transformation, the original SARs and EARs collapsed into an approximately single curve (Fig. 2a, b). This collapse was also observed by using an alternative sampling design based on continental (self-similar) instead of quadratic shapes of study plots (Fig. 2c, d; see Methods). The steeper SARs observed for amphibians (Fig. 1a) can thus be attributed to their considerably smaller ranges: the SAR increases rapidly at smaller absolute areas, and the slope continues to increase, whereas the other two taxa with much larger range sizes never approach a similarly steep relationship (see Supplementary Fig. 11).

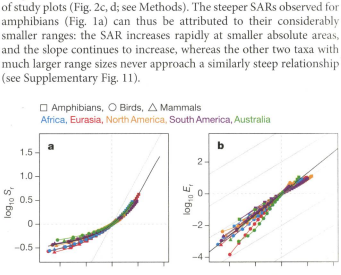


Figure 2 | SARs and EARs after rescaling. a, **b**, After expressing the area in units corresponding to mean range size and standardizing the vertical axis so that it represents species richness relative to mean richness for a given unit area, all the SARs and EARs collapse into one universal relationship, although some deviations exist, particularly in small areas. For EARs (**b**), birds in Eurasia and Australia represent the only considerable deviations. In these regions many endemics with small ranges occur at the edge of the continents, whereas the areas for which the EARs were calculated were taken predominantly from the centre of the continent. **c**, **d**, These universal relationships also exist for SARs and EARs constructed using the alternative, continental shape design, in which sample areas are not quadrats but keep the shape of the given continent (see Methods and Supplementary Discussion). Solid black lines refer to rescaled SARs and EARs predicted by simulations based on a random placement of simplified ranges (model 3; see Fig. 3 and Methods). Solid grey lines all have slope of 1. For explanations of A_r, S_r and E_r see equations (1), (2) and (3).

To test whether the observed continental relationships and their collapse can be predicted on the basis of simple assumptions concerning spatial distribution of species ranges, we developed several spatially explicit simulation models (see Methods). Specifically, we assessed the degree to which the observed SARs and EARs may be recovered by using placement of species ranges modelled as simple contiguous shapes. Although multiple evolutionary and ecological factors ultimately determine the exact location and sizes of geographic ranges (which in turn affect SAR and EAR), we find that just a few assumptions are sufficient to explain the observed scaling relationships (Fig. 3). Among our models those with independent (models 3 and 4; see Fig. 3 and Methods for details), as opposed to clumped (models 1 and 2), placement of geographic ranges produce SAR and EAR shapes as well as collapses of all curves that are essentially identical to those observed. The resulting patterns are not particularly sensitive to the exact shape of the frequency distribution of range sizes, because model 3 (which retained the observed range size distributions) provided very similar patterns to those of model 4, which did not retain them (see Supplementary Discussion). In contrast with models 3 and 4, the observed geographic range locations and sizes do show spatial non-independence, but much less so than in model 2, and the effect on the observed collapse is minimal (Supplementary Figs 13–15 and Supplementary Discussion). The empirical patterns (Fig. 2) are therefore expected whenever a species distribution is represented by more or less independently located contiguous ranges, with the mean range size of species being the only biologically relevant variable affecting the exact properties of the patterns.

The universality of SARs and EARs after rescaling implies that a knowledge of mean species richness (of either endemics or all species) at one scale allows the estimation of the whole SARs and EARs with only one additional piece of information: mean range size for species in the region. Although this information may not usually be available without knowledge of the geographic distribution of all species (and thus also the whole SARs and EARs), in some cases it may be estimated

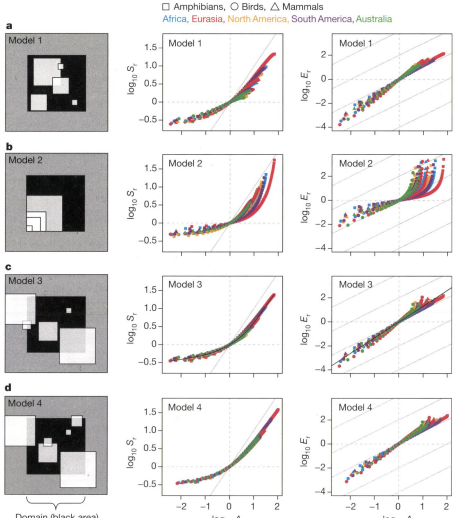


Figure 3 | Rescaled SARs and EARs predicted by four simulation models of range placement. Range sizes were drawn from the empirical frequency distributions of each taxon and domain (black areas in Supplementary Fig. 3) and were placed into a domain with a size equal to that of the regions analysed for the original SARs and EARs using the strictly nested quadrat design (Supplementary Fig. 3 and Methods; for the results based on the regions analysed using the continent shape design see Supplementary Fig. 12). **a**, Model 1 is based on a random placement of square ranges within the domain, producing a higher concentration of range midpoints in the centre of the domain (mid-domain effect[30]). **b**, Model 2 places all the square ranges in a corner of the domain, to illustrate the role of non-random range position. **c**, Model 3 is based on a random placement of square ranges but minimizes the mid-domain effect by allowing model ranges to overlap the domain only partly. The observed frequency distribution of range sizes is retained, but resulting range shapes within the domain become variable. **d**, Model 4 is similar to model 3 and completely avoids the mid-domain effect but does not retain the originally observed range size distribution (see Methods for details). We produce a fitted line for model 3 results to highlight its match with the empirical patterns (see Fig. 2): black lines represent the Lowess regression line for the rescaled SAR plot (smoothing span 0.2) and the linear regression line for the rescaled EAR plot. Solid grey lines all have slope of 1.

reasonably well from similar taxa or representative subtaxa. However, the larger deviations from the universal relationship in small areas (Fig. 2a) will probably limit the accuracy of richness estimates towards smaller spatial scales. This is expected, because the SAR and EAR at these scales are determined by complex shapes of individual geographic ranges[22], which go beyond simple differences in range sizes.

In contrast to the strongly nonlinear EARs previously reported for small spatial scales[9], the slope of the continental EARs assessed here was generally close to 1. Therefore, for the scales examined, the number of species predicted to go extinct is roughly proportional to the area destroyed. However, the uncertainty in any such predictions would be large (Table 1), and the shape of the EAR is unlikely to hold up at the finer spatial scales relevant to conservation[9]. In addition, the relationships addressed here comprise only the direct effect of shrinking habitable area and not the cascading effects of species interactions or other consequences of species loss. Nevertheless, for the scales analysed here, the relatively steep slopes of all EARs suggest large extinction rates from area loss.

Spatial biodiversity patterns, including the SAR and EAR, are affected by many factors, ranging from the spatial arrangements of continental masses and biomes and the patterns of diversification and dispersal playing out within and among them, to population dynamics and interspecific interactions[23,24]. However, all of these processes ultimately translate into patterns of geographic distribution of individual species, which then represent a proximate driver of spatial macroecological patterns[22,25]. We have shown that species range sizes have a key role in large-scale upward-accelerating SARs, and consequently also EARs at specific spatial scales. Because local SAR slopes are mathematically related to species spatial turnover[14,26,27], the determinants of species range sizes predictably determine global patterns in species spatial turnover, the SAR and the EAR. These findings suggest that an integrated evolutionary and ecological understanding of just a few attributes of regional biota can enable far-reaching predictions to be made about the scaling of biodiversity.

METHODS SUMMARY

We calculated mean species richness across all possible quadrats of given size by using the strictly nested quadrat method[28,29]. We calculated SARs for continental regions able to accommodate quadrats encompassing 20×20 grid cells, except for Australia where because of the smaller size we used 14×14 (Supplementary Fig. 3). This avoids potential biases resulting from different species richness in marginal areas that could not be sampled by large quadrats (see Supplementary Fig. 4). In a second sampling design we adjusted plot boundaries to mimic continental shapes (Supplementary Fig. 5), which increased the amount of edge area that could be included in regions such as North America with more complicated geometry. This procedure yielded qualitatively similar, but noisier, results (see Methods, Supplementary Discussion, Supplementary Fig. 6 and Supplementary Table 2 for details). We used four simulation models of range placement within a domain to examine the effects of range position (random or spatially clumped) and shape (constant or varying) on resulting SARs and EARs (see Fig. 3 for more details on the models).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 5 December 2011; accepted 10 May 2012.
Published online 24 June 2012.

1. Rosenzweig, M. L. *Species Diversity in Space and Time* (Cambridge Univ. Press, 1995).
2. Allen, A. P. & White, E. P. Effects of range size on species–area relationships. *Evol. Ecol. Res.* **5,** 493–499 (2003).
3. McGill, B. & Collins, C. A unified theory for macroecology based on spatial patterns of abundance. *Evol. Ecol. Res.* **5,** 469–492 (2003).
4. Hubbell, S. P. *The Unified Theory of Biodiversity and Biogeography* (Princeton Univ. Press, 2001).
5. Rosindell, J. & Cornell, S. J. Species–area relationships from a spatially explicit neutral model in an infinite landscape. *Ecol. Lett.* **10,** 586–595 (2007).
6. O'Dwyer, J. P. & Green, J. L. Field theory for biogeography: a spatially explicit model for predicting patterns of biodiversity. *Ecol. Lett.* **13,** 87–95 (2010).
7. Lawton, J. H. & May, R. (eds) *Extinction Rates* (Oxford Univ. Press, 1995).
8. Pimm, S. L. & Raven, P. Biodiversity: extinction by numbers. *Nature* **403,** 843–845 (2000).
9. He, F. & Hubbell, S. P. Species–area relationships always overestimate extinction rates from habitat loss. *Nature* **473,** 368–371 (2011).
10. Smith, A. B. Caution with curves: caveats for using the species–area relationship in conservation. *Biol. Conserv.* **143,** 555–564 (2010).
11. Drakare, S., Lennon, J. J. & Hillebrand, H. The imprint of the geographical, evolutionary and ecological context on species–area relationships. *Ecol. Lett.* **9,** 215–227 (2006).
12. Connor, E. F. & McCoy, E. D. The statistics and biology of the species–area relationship. *Am. Nat.* **113,** 791–833 (1979).
13. Harte, J., Smith, A. B. & Storch, D. Biodiversity scales from plots to biomes with a universal species–area curve. *Ecol. Lett.* **12,** 789–797 (2009).
14. Šizling, A. L., Kunin, W. E., Šizlingová, E., Reif, J. & Storch, D. Between geometry and biology: the problem of universality of the species–area relationship. *Am. Nat.* **178,** 602–611 (2011).
15. Kinzig, A. P. & Harte, J. Implications of endemics–area relationships for estimates of species extinctions. *Ecology* **81,** 3305–3311 (2000).
16. Pereira, H. M., Borda-de-Água, L. & Martins, I. S. Geometry and scale in species–area relationships. *Nature* **482,** E3–E4 (2012).
17. Green, J. L. & Ostling, A. Endemics–area relationships: the influence of species dominance and spatial aggregation. *Ecology* **84,** 3090–3097 (2003).
18. Belmaker, J. & Jetz, W. Cross-scale variation in species richness–environment associations. *Glob. Ecol. Biogeogr.* **20,** 464–474 (2011).
19. Hurlbert, A. H. & Jetz, W. Species richness, hotspots, and the scale dependence of range maps in ecology and conservation. *Proc. Natl Acad. Sci. USA* **104,** 13384–13389 (2007).
20. Storch, D. *et al.* Energy, range dynamics and global species richness patterns: reconciling mid-domain effects and environmental determinants of avian diversity. *Ecol. Lett.* **9,** 1308–1320 (2006).
21. Buckley, L. B. & Jetz, W. Environmental and historical constraints on global patterns of amphibian richness. *Proc. R. Soc. Lond. B* **274,** 1167–1173 (2007).
22. Storch, D. *et al.* The quest for a null model for macroecological patterns: geometry of species distributions at multiple spatial scales. *Ecol. Lett.* **11,** 771–784 (2008).
23. Losos, J. B. & Schluter, D. Analysis of an evolutionary species–area relationship. *Nature* **408,** 847–850 (2000).
24. Ricklefs, R. E. A comprehensive framework for global patterns in biodiversity. *Ecol. Lett.* **7,** 1–15 (2004).
25. McGill, B. J. Towards a unification of unified theories of biodiversity. *Ecol. Lett.* **13,** 627–642 (2010).
26. Harte, J. & Kinzig, A. P. On the implications of species–area relationships for endemism, spatial turnover, and food web patterns. *Oikos* **80,** 417–427 (1997).
27. Arita, H. T. & Rodríguez, P. Geographic range, turnover rate and the scaling of species diversity. *Ecography* **25,** 541–550 (2002).
28. Scheiner, S. M. Six types of species–area curves. *Glob. Ecol. Biogeogr.* **12,** 441–447 (2003).
29. Leitner, W. A. & Rosenzweig, M. L. Nested species–area curves and stochastic sampling: a new theory. *Oikos* **79,** 503–512 (1997).
30. Colwell, R. K. & Lees, D. C. The mid-domain effect: geometric constraints on the geography of species richness. *Trends Ecol. Evol.* **15,** 70–76 (2000).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank J. Belmaker, K. Mertes-Schwartz, C. Sheard and D. Rosauer for useful comments. The study was supported by the Grant Agency of the Czech Republic (P505/11/2387), the Czech Ministry of Education (MSM0021620845) and the EU FP7 SCALES project ('Securing the Conservation of biodiversity across Administrative Levels and spatial, temporal and Ecological Scales'; project No. 26852). W.J. acknowledges support from National Science Foundation grants DBI 0960550 and DEB 1026764, and NASA Biodiversity Program grant number NNX11AP72G.

Author Contributions D.S. initiated the research. D.S., P.K. and W.J. developed the ideas, methods and concepts, and wrote the manuscript. W.J. adjusted and provided the data. P.K. performed the analyses and simulations.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to D.S. (storch@cts.cuni.cz).

LETTER

doi:10.1038/nature11281

A complete insect from the Late Devonian period

Romain Garrouste[1], Gaël Clément[2], Patricia Nel[1], Michael S. Engel[3], Philippe Grandcolas[1], Cyrille D'Haese[1], Linda Lagebro[4], Julien Denayer[5], Pierre Gueriau[2,6], Patrick Lafaite[7], Sébastien Olive[5,8], Cyrille Prestianni[8] & André Nel[1]

After terrestrialization, the diversification of arthropods and vertebrates is thought to have occurred in two distinct phases[1], the first between the Silurian and the Frasnian stages (Late Devonian period) (425–385 million years (Myr) ago), and the second characterized by the emergence of numerous new major taxa, during the Late Carboniferous period (after 345 Myr ago). These two diversification periods bracket the depauperate vertebrate Romer's gap (360–345 Myr ago) and arthropod gap (385–325 Myr ago)[1], which could be due to preservational artefact[2,3]. Although a recent molecular dating has given an age of 390 Myr for the Holometabola[4], the record of hexapods during the Early–Middle Devonian (411.5–391 Myr ago, Pragian to Givetian stages) is exceptionally sparse and based on fragmentary remains, which hinders the timing of this diversification. Indeed, although Devonian Archaeognatha are problematic[5,6], the Pragian of Scotland has given some Collembola and the incomplete insect *Rhyniognatha*, with its diagnostic dicondylic, metapterygotan mandibles[5,7]. The oldest, definitively winged insects are from the Serpukhovian stage (latest Early Carboniferous period)[8]. Here we report the first complete Late Devonian insect, which was probably a terrestrial species. Its 'orthopteroid' mandibles are of an omnivorous type, clearly not modified for a solely carnivorous diet. This discovery narrows the 45-Myr gap in the fossil record of Hexapoda, and demonstrates further a first Devonian phase of diversification for the Hexapoda, as in vertebrates, and suggests that the Pterygota diversified before and during Romer's gap.

The insect was recovered from the Famennian Strud locality, Namur province, Belgium (50° 26′ 43.32″ N, 5° 03′ 24.86″ E), known widely for its early tetrapods[9,10]. It was found in a freshwater association including plants[11], numerous Crustacea (Branchiopoda and Malacostraca) and Chelicerata (Eurypterida) (Fig. 1, Supplementary Figs 1 and 2 and Supplementary Information 1). The specimen is comprised of a two-dimensional compression with median abdominal structures elevated owing to natural filling of the gut, and excludes the possibility that the material represents an exuvia. A 'shadow' of organic origin surrounds the body. The connections of the appendages with the body are partly destroyed because of a well-known compression and decay process[12], rendering difficult the study of some parts. The appendages were not displaced, except for one or two legs.

Class Insecta
Clade Dicondylia
Strudiella devonica gen. et sp. nov.

Etymology. *Strudiella* is a diminutive form based on the type locality Strud (the name is feminine); *devonica* is after the Devonian age of the fossil.

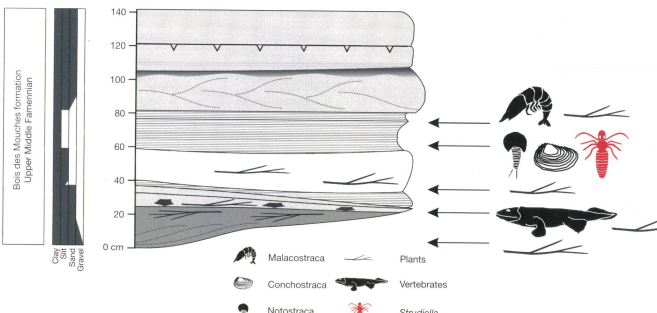


Figure 1 | Partial Strud stratigraphy (fossiliferous levels), Bois des Mouches Formation, Upper Famennian.

[1]UMR CNRS 7205, CP 50, Entomologie, Muséum national d'Histoire naturelle, 45 rue Buffon, F-75005 Paris, France. [2]UMR CNRS 7207, Paléontologie, Muséum national d'Histoire naturelle, 8 rue Buffon, F-75005 Paris, France. [3]Division of Entomology, Natural History Museum, and Department of Ecology and Evolutionary Biology, 1501 Crestline Drive – Suite 140, University of Kansas, Lawrence, Kansas 66045, USA. [4]Department of Earth Sciences, Uppsala University, Villavägen 16, SE-752 36 Uppsala, Sweden. [5]Service de Paléontologie animale et humaine, Département de Géologie, Université de Liège, Bât. B.18, Allée du Six-Août, Sart Tilman, B-4000 Liège, Belgium. [6]IPANEMA, USR 3461 CNRS–Ministère de la Culture et de la Communication, F-91190, Saint-Aubin, France. [7]CNRS–MNHN DICAP, Service Multimédia, CP 27, 57 rue Cuvier, F-75005 Paris, France. [8]Royal Belgian Institute of Natural Sciences Paleontology Department, 29, Rue Vautier, B-1000 Brussels, Belgium.

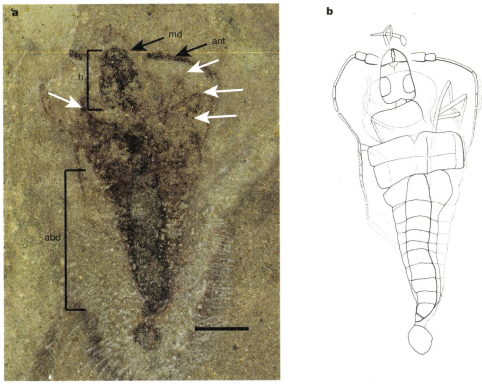


Figure 2 | General habitus of *Strudiella devonica* gen. et sp. nov. a, Photograph of the part. **b**, Reconstruction of general habitus. Scale bar, 1 mm. White arrows indicate legs visible on part. abd, abdomen; ant, antenna; h, head; md, mandible.

Holotype. IRSNB a12818a–b, part and counterpart, by present designation.

Diagnosis and description. Apterous, body elongate and narrow, 8.0 mm long and 1.7 mm wide (Fig. 2a, b); six thoracic, uniramous legs with tibiae and femora long and thin; antenna uniramous (Figs 2a and 3a), long, with scape and pedicel distinctly broader than remaining antennomeres, about 10 short flagellomeres; mandible triangular (of metapterygotan form) with a continuous series of sharp but small irregular molar and incisor cusps (Fig. 3b, c and Supplementary Fig. 3); large dark eyes in posterior part of head; head rather small; thorax broad, well separated from head and abdomen, with a rounded structure covering posterior half of head, corresponding to an expanded pronotum; abdomen divided into 10 segments, without lateral leglets, gills or other appendicular structures.

The presence of a thorax separated from the head and abdomen, bearing three pairs of legs, is one of the hallmark apomorphies of the Hexapoda[13]. Further features of significance are long, uniramous legs, one pair of long, uniramous antennae, large eyes, abdomen divided into 10 segments and absence of abdominal leglets. Although these can be found individually in different crustacean clades[14,15], their combination with the aforementioned apomorphy of hexapods is distinctly insectan and precludes an interpretation of this fossil as a juvenile Notostraca, which are abundant in the layer (see Supplementary Information). The scape and pedicel being distinctly broader than

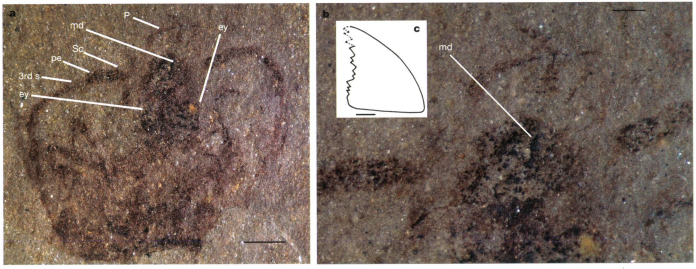


Figure 3 | *Strudiella devonica* gen. et sp. nov., counterpart details. **a**, Photograph of anterior part of head. **b**, Photograph of left mandible. **c**, Reconstruction of left mandible. Scale bars, 0.45 mm (**a**), 0.2 mm (**b**), 0.1 mm (**c**). 3rd S, third antennal segment; ey, eye; md, mandible; p, maxillary palp; pe, pedicel; Sc, scape.

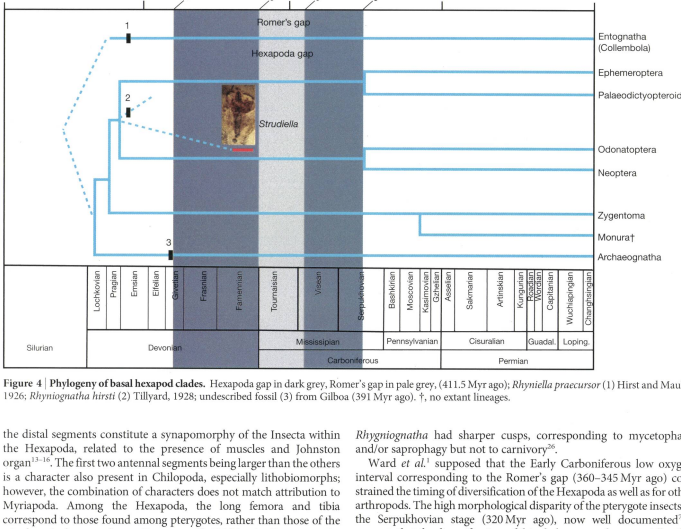


Figure 4 | Phylogeny of basal hexapod clades. Hexapoda gap in dark grey, Romer's gap in pale grey, (411.5 Myr ago); *Rhyniella praecursor* (1) Hirst and Maulik, 1926; *Rhyniognatha hirsti* (2) Tillyard, 1928; undescribed fossil (3) from Gilboa (391 Myr ago). †, no extant lineages.

the distal segments constitute a synapomorphy of the Insecta within the Hexapoda, related to the presence of muscles and Johnston organ[13–16]. The first two antennal segments being larger than the others is a character also present in Chilopoda, especially lithobiomorphs; however, the combination of characters does not match attribution to Myriapoda. Among the Hexapoda, the long femora and tibia correspond to those found among pterygotes, rather than those of the wingless Archaeognatha or Zygentoma[17]. Within the Pterygota, the mandibles are of an 'orthopteroid' type, short and triangular[6,18–22] and currently considered an apomorphy of the winged Metapterygota (Pterygota, excluding Palaeodictyoptera and Ephemeroptera)[5,6,18,19]. Unfortunately, wings are not observable on the present specimen. The absence of wings plus the minute size suggests that the individual was a nymph, but a conclusive determination about the absence of genital structures cannot be established owing to the poor preservation of the abdominal apex.

The expanded pronotum is similar to those of many insect lineages (for example, Dictyoptera and Grylloblattodea), and its presence in a Middle-Palaeozoic-era insect is not surprising. The lack of lateral appendages (leglets and gills) on the abdominal segments is clearly derived relative to the condition in the wingless Zygentoma, and differs from nymphs (but not adults) of several pterygote orders, although whether these are individually derived in those immatures or are plesiomorphic is debatable[23]. *Strudiella* shares, with the most basal hexapod lineages, antennae with uniformly similar flagellomeres[5,15,24].

The long legs without adaptations for swimming, plus the apparent absence of abdominal gills and rarity of *Strudiella* within the arthropod fauna of the Strud locality, support the hypothesis that it was a terrestrial animal. It must have been omnivorous or phytophagous but certainly not carnivorous given the weak development of the incisor compared with the molar cusps, and the sharp irregular cusps corresponding to the 'omnivorous type' of Gangwere[25]. *Strudiella* and *Rhyniognatha* certainly had different feeding habits as the mandibles of *Rhyniognatha* had sharper cusps, corresponding to mycetophagy and/or saprophagy but not to carnivory[26].

Ward *et al.*[1] supposed that the Early Carboniferous low oxygen interval corresponding to the Romer's gap (360–345 Myr ago) constrained the timing of diversification of the Hexapoda as well as for other arthropods. The high morphological disparity of the pterygote insects in the Serpukhovian stage (320 Myr ago), now well documented[7,27], suggests that the diversification of this clade occurred very rapidly after Romer's gap, or more likely during it, in accordance with the recent discoveries of early Carboniferous terrestrial arthropods[2]. *Strudiella* demonstrates further that an early diversification of the dicondylic insects occurred before Romer's gap[4] (Fig. 4), well in accordance with the presence of a diversified abundant terrestrial vegetation, including forests, since the mid-Devonian[6,11,28–30].

METHODS SUMMARY

The material is housed at the Royal Belgian Institute of Natural Sciences (Brussels, Belgium). The fossils were prepared using a sharp knife. Photographs were taken using an Olympus SZX9 stereomicroscope system with an Olympus E3 digital camera, and the fossil was moistened with 70% alcohol. Illustrations were prepared using a camera lucida on a binocula Olympus SZX9.

Received 3 April 2012; accepted 1 June 2012.

1. Ward, P., Labandeira, C., Laurin, M. & Berner, R. A. Confirmation of Romer's gap as a low oxygen interval constraining the timing of initial arthropod and vertebrate terrestrialization. *Proc. Natl Acad. Sci. USA* **103,** 16818–16822 (2006).
2. Smithson, T. R., Wood, S. P., Marshall, J. E. A. & Clack, J. A. Earliest Carboniferous tetrapod and arthropod faunas from Scotland populate Romer's gap. *Proc. Natl Acad. Sci. USA* **109,** 4532–4537 (2012).
3. Retallack, G. J. Woodland hypothesis for Devonian evolution of tetrapods. *J. Geol.* **119,** 235–258 (2011).
4. Rehm, P. *et al.* Dating the arthropod tree based on large-scale transcriptome data. *Mol. Phyl. Evol.* **61,** 880–887 (2011).
5. Grimaldi, D. & Engel, M. S. *Evolution of the Insects* (Cambridge Univ. Press, 2005).
6. Shear, W. A. & Selden, P. A. in *Plants Invade the Land. Evolutionary & Environmental Perspectives* (eds Gensel, P. G. & Edwards, D.) 29–51 (Columbia Univ. Press, 2001).
7. Engel, M. S. & Grimaldi, D. New light shed on the oldest insect. *Nature* **427,** 627–630 (2004).

8. Prokop, J., Nel, A. & Hoch, I. Discovery of the oldest known Pterygota in the Lower Carboniferous of the Upper Silesian Basin in the Czech Republic (Insecta: Archaeorthoptera). *Geobios* **38,** 383–387 (2005).
9. Clément, G. *et al.* Devonian tetrapod from Western Europe. *Nature* **427,** 412–413 (2004).
10. Blieck, A., Clément, G. & Streel, M. in *The Terrestrialization Process: Modelling Complex Interactions at the Biosphere–Geosphere Interface* (eds Vecoli, M., Clément, G. & Meyer-Berthaud, B.) 129–138 (Geological Society of London, 2010).
11. Prestianni, C., Streel, M., Thorez, J. & Gerrienne, P. Strud: old quarry, new discoveries. Preliminary report. *Carnets Géol. Mémoir.* **2007,** 43–47 (2007).
12. Martínez-Delclòs, X., Briggs, D. E. G. & Peñalver, E. Taphonomy of insects in carbonates and amber. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **3225,** 1–46 (2004).
13. Bitsch, C. & Bitsch, J. Phylogenetic relationships of basal hexapods among the mandibulate arthropods: a cladistic analysis based on comparative morphological characters. *Zool. Scr.* **33,** 511–550 (2004).
14. McLaughlin, P. A. *Comparative Morphology of Recent Crustacea* (W. H. Freeman, 1980).
15. Imms, A. D. On the antennal musculature in insects and other arthropods. *Q. J. Microsc. Sci.* **81,** 273–320 (1939).
16. Yack, J. E. The structure and function of auditory chordotonal organs in insects. *Microsc. Res. Tech.* **63,** 315–337 (2004).
17. Sturm, H. & Machida, R. *Archaeognatha. Handbook of Zoology*, Vol. 4 of *Arthropoda: Insecta*, i–viii, 1–213 (Berlin: Walter de Gruyter, 2001).
18. Furst Von Lieven, A. The transformation from monocondylous to dicondylous mandibles in the Insecta. *Zool. Anz.* **239,** 139–146 (2000).
19. Kukalová-Peck, J. Phylogeny of higher taxa in Insecta: finding synapomorphies in the extant fauna and separating them from homoplasies. *Evol. Biol.* **35,** 4–51 (2008).
20. Staniczek, A. H. The mandible of silverfish (Insecta: Zygentoma) and mayflies (Ephemeroptera): its morphology and phylogenetic significance. *Zool. Anz.* **239,** 147–178 (2000).
21. Bitsch, J. The arthropod mandible: morphology and evolution. Phylogenetic implications. *Ann. Soc. Entomol. Fr. (N.S.)* **37,** 305–321 (2001).
22. Snodgrass, R. E. Comparative studies on the jaws of mandibulate arthropods. *Smithson. Misc. Coll.* **116,** 1–85 (1952).
23. Kukalová-Peck, J. Carboniferous protodonatoid dragonfly nymphs and the synapomorphies of Odonatoptera and Ephemeroptera (Insecta: Palaeoptera). *Palaeodiversity* **2,** 169–198 (2009).
24. Bechly, G., Brauckmann, C., Zessin, W. & Gröning, E. New results concerning the morphology of the most ancient dragonflies (Insecta: Odonatoptera) from the Namurian of Hagen-Vorhalle (Germany). *Z. Zool. Syst. Evol.* **39,** 209–226 (2001).
25. Gangwere, S. K. The structural adaptations of mouthparts in Orthoptera and allies. *Eos. Rev. Esp. Entomol.* **41,** 67–85 (1965).
26. Guthrie, D. M. & Tindall, A. R. *The Biology of the Cockroach* (Edward Arnold, 1968).
27. Brauckmann, C., Schöllmann, L. & Sippel, W. Die fossilen Insekten, Spinnentiere und Eurypteriden von Hagen-Vorhalle. *Geol. Paläontol. Westfalen* **59,** 1–89 (2003).
28. Labandeira, C. C. Silurian to Triassic plant and hexapod clades and their associations: new data, a review, and interpretations. *Arthr. Syst. Phyl.* **64,** 53–94 (2006).
29. Labandeira, C. C. The origin of herbivory on land: initial patterns of plant tissue consumption by arthropods. *Insect Sci.* **14,** 259–275 (2007).
30. Stein, W. E., Berry, C. M., VanAller Hernick, L. & Mannolini, F. Surprisingly complex community discovered in the mid-Devonian fossil forest at Gilboa. *Nature* **483,** 78–81 (2012).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank O. Béthoux who discovered and prepared most of the arthropod material from Strud including the specimen described herein. We also thank G. Budd and G. Edgecombe for discussion on the fossil material and improving the first version of the paper, Gesves local council staff and field workers of the Strud expeditions, G. Odebert and S. Fernandez for preparing illustrations, and C. Lemzaouda and O. Béthoux for photographs of the associated arthropod fauna. Thanks are due to A. Folie for our request of a catalogue number for the specimen described herein (requests for materials can be sent to afolie@naturalsciences.be). This work was partly supported by the French National Agency under the TERRES project (number ANR-2010-BLAN-607). Support for M.S.E. was provided by US National Science Foundation grant DEB-0542909.

Author Contributions R.G., P.N. and G.C. are first authors with equal rank; R.G., A.N., P.N., P.G., C.D'H., L.L., M.S.E., J.D., C.P., P.G. and S.O. drafted the manuscript and prepared figures. A.N. and P.N. coordinated the manuscript; G.C. coordinated and participated in fieldwork at the Strud locality and contributed to the draft manuscript; L.L., J.D., C.P., P.G. and S.O. also participated in fieldwork.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to R.G. (garroust@mnhn.fr) or A.N. (anel@mnhn.fr).

LETTER

doi:10.1038/nature11237

Defining the core *Arabidopsis thaliana* root microbiome

Derek S. Lundberg[1,2]*, Sarah L. Lebeis[1]*, Sur Herrera Paredes[1]*, Scott Yourstone[1,3]*, Jase Gehring[1], Stephanie Malfatti[4], Julien Tremblay[4], Anna Engelbrektson[4]†, Victor Kunin[4]†, Tijana Glavina del Rio[4], Robert C. Edgar[5], Thilo Eickhorst[6], Ruth E. Ley[7], Philip Hugenholtz[4,8], Susannah Green Tringe[4] & Jeffery L. Dangl[1,2,9,10,11]

Land plants associate with a root microbiota distinct from the complex microbial community present in surrounding soil. The microbiota colonizing the rhizosphere (immediately surrounding the root) and the endophytic compartment (within the root) contribute to plant growth, productivity, carbon sequestration and phytoremediation[1–3]. Colonization of the root occurs despite a sophisticated plant immune system[4,5], suggesting finely tuned discrimination of mutualists and commensals from pathogens. Genetic principles governing the derivation of host-specific endophyte communities from soil communities are poorly understood. Here we report the pyrosequencing of the bacterial 16S ribosomal RNA gene of more than 600 *Arabidopsis thaliana* plants to test the hypotheses that the root rhizosphere and endophytic compartment microbiota of plants grown under controlled conditions in natural soils are sufficiently dependent on the host to remain consistent across different soil types and developmental stages, and sufficiently dependent on host genotype to vary between inbred *Arabidopsis* accessions. We describe different bacterial communities in two geochemically distinct bulk soils and in rhizosphere and endophytic compartments prepared from roots grown in these soils. The communities in each compartment are strongly influenced by soil type. Endophytic compartments from both soils feature overlapping, low-complexity communities that are markedly enriched in Actinobacteria and specific families from other phyla, notably Proteobacteria. Some bacteria vary quantitatively between plants of different developmental stage and genotype. Our rigorous definition of an endophytic compartment microbiome should facilitate controlled dissection of plant–microbe interactions derived from complex soil communities.

Roots influence the rhizosphere by altering soil pH, soil structure, oxygen availability, antimicrobial concentration, and quorum-sensing mimicry, and by providing an energy source of dead root material and carbon-rich exudates[6,7]. The microbiota inhabiting this niche can both benefit and undermine plant health; shifting this balance is of agronomic interest. Mutualistic microbes may provide the plant with physiologically accessible nutrients and phytohormones that improve plant growth, may suppress phytopathogens or may help plants withstand heat, salt and drought[8,9]. The rhizosphere community is a subset of soil microbes that are subsequently filtered via niche utilization attributes and interactions with the host to inhabit the endophytic compartment[10] (EC). Although a variety of microbes may enter and become transient endophytes, those consistently found inside roots are candidate symbionts or stealthy pathogens[10,11]. Notably, *Arabidopsis* and other Brassicaceae are not well colonized by arbuscular mycorrhizal fungi, implying that other microorganisms may fill this niche.

Microbial community structure differs across plant species[12,13], and there are reports of host-genotype-dependent differences in patterns of microbial associations[14,15]. However, the divergent methods used in those studies relied on small sample sizes and low-resolution phylotyping techniques potentially confounded by off-target sequences and chimaeric amplicons. We developed a robust experimental system to sample repeatedly the root microbiome using high-throughput sequencing. Our results confirm many of the general conclusions from earlier studies and, because of controlled experimental design and the power of deep sequencing, provide a key step towards the definition of this microbiome's functional capacity and the host genes that potentially contribute to microbial association phenotypes. Such plant genes would constitute major agronomic targets.

We used 454 pyrosequencing to sequence 16S ribosomal RNA (rRNA) gene amplicons for DNA prepared from eight diverse, inbred *A. thaliana* accessions. Plants were grown from surface-sterile seeds in climate-controlled conditions in two diverse soils, respectively termed Mason Farm and Clayton (Supplementary Table 1; detailed in Supplementary Information). For each soil, we assayed multiple individuals from each *A. thaliana* accession grown from sterile seeds in both soils across independent full-factorial biological replicates, in which all genotypes and bulk soils (pots without a plant) for a given soil type were grown in parallel (Supplementary Table 2). We isolated separate rhizosphere and EC fractions from individual plant root systems (Supplementary Fig. 1 and Supplementary Table 2). We established 1114F and 1392R as our primer pair (Supplementary Information and Supplementary Fig. 2). Using an otupipe-based pipeline (http://drive5.com/otupipe/), we grouped sequences into 97%-identical operational taxonomic units (OTUs), reduced noise and removed chimaeras. We determined technical reproducibility thresholds to conclude that OTUs defined by ≥25 reads in ≥5 samples (hereafter 25 × 5) are individually 'measurable OTUs'[16,17] (Supplementary Figs 2 and 10). All data reported here are from one run of our otupipe-based pipeline (Supplementary Fig. 3 and Supplementary Database 1).

Excluding additional control samples, we ribotyped 1,248 samples comprising 111 bulk soil, 613 rhizosphere and 524 EC samples, generating 9,787,070 high-quality reads (Supplementary Figs 3 and 4a–c). After removing plant-sequence-derived OTUs, we obtained a table of usable OTU read counts per sample containing 6,387,407 reads distributed across 18,783 OTUs. We normalized this table of usable reads by rarefying to 1,000 reads per sample (Supplementary Database 2a) or, alternatively, by dividing the reads per OTU in a sample by the sum of usable reads in that sample, resulting in a table

[1]Department of Biology, University of North Carolina, Chapel Hill, North Carolina 27599, USA. [2]Curriculum in Genetics and Molecular Biology, University of North Carolina, Chapel Hill, North Carolina 27599, USA. [3]Curriculum in Bioinformatics and Computational Biology, University of North Carolina, Chapel Hill, North Carolina 27599, USA. [4]DOE Joint Genome Institute, Walnut Creek, California 94598, USA. [5]Taxon Biosciences, Inc., Tiburon, California 94920, USA. [6]Soil Science, Faculty of Biology and Chemistry, University of Bremen, Bremen 28359, Germany. [7]Department of Microbiology, Cornell University, Ithaca, New York 14853, USA. [8]Australian Centre for Ecogenomics, School of Chemistry and Molecular Biosciences & Institute for Molecular Bioscience, The University of Queensland, Brisbane, Queensland 4072, Australia. [9]Department of Microbiology and Immunology, University of North Carolina, Chapel Hill, North Carolina 27599, USA. [10]Carolina Center for Genome Sciences, University of North Carolina, Chapel Hill, North Carolina 27599, USA. [11]Howard Hughes Medical Institute, University of North Carolina, Chapel Hill, North Carolina 27599, USA. †Present addresses: Department of Plant and Microbial Biology, University of California, Berkeley, California 94720-3102, USA (A.E.); Taxon Biosciences, Inc., Tiburon, California 94920, USA (V.K.).
*These authors contributed equally to this work.

of relative abundances (frequencies) (Supplementary Database 2b). Using the 25×5 threshold, we defined 778 measurable OTUs representing 54% (3,463,632) of the usable reads (Supplementary Fig. 4c and Supplementary Table 3). The diversity of the 778 measurable OTUs in soil, rhizosphere and EC fractions showed expected relative trends when compared with the diversity by fraction of all usable OTUs (Supplementary Fig. 4d). We display the rarefaction-normalized data; parallel analyses of frequency-normalized data are provided in Supplementary Figures.

We used principal coordinate analysis on pairwise, normalized, weighted UniFrac distances between all samples, considering all usable OTUs, to identify the main factors driving community composition (Fig. 1a and Supplementary Fig. 5a). The first principal coordinate (PCo1) revealed that the two bulk soils and their associated rhizospheres were differentiated from the respective EC fractions. Soil type was the main factor in the second component (PCo2). This pattern was recapitulated by hierarchical clustering of pairwise Bray–Curtis dissimilarities considering only measurable OTUs (Fig. 1b and Supplementary Fig. 5b). Samples harvested at different developmental stages clustered together, indicating that this variable does not have a major effect on overall community composition (Fig. 1 and Supplementary Fig. 5a, b; yng versus old, where yng refers to the time of appearance of an inflorescence meristem and old refers to fruiting plants with greater than 50% senescent leaves). Additional control samples from the reference genotype Col-0 harvested from four independent digs of Mason Farm soil underscored the reproducibility of these bacterial community profiles (Supplementary Fig. 6). Together, these data demonstrate that the interaction of diverse soil communities with plants determines the assembly of the rhizosphere, leading to winnowed ECs, that the ECs from at least these two diverse soils are very different from the starting soil communities and that there is little difference in communities over host developmental time.

We fitted a general linear mixed model (GLMM) to samples from each set of plant fractions (rhizosphere or EC), plus the bulk soil controls, to identify measurable OTUs whose abundances differ significantly between plant and bulk soil as a result of soil type, developmental stage, fraction and genotype (Supplementary Information and Supplementary Database 3). This approach allowed us to quantify the contribution from each variable to the community composition (Supplementary Table 4). Controlling for sequencing plate effects, plant fraction is the most important factor; its effect is strongest for the EC, consistent with our UniFrac and Bray–Curtis analyses. Soil type is less important, followed by experiment, developmental stage and, finally, genotype, which had a small but consistent effect.

Hierarchical clustering of sample groups considering 256 OTUs identified by the GLMM to differentiate rhizosphere and EC from soil recapitulated the separation of EC from soil and rhizosphere (Fig. 2A and Supplementary Fig. 7a, left; compare with Fig. 1 and Supplementary Fig. 5). Of these, 164 OTUs were enriched in EC samples (Fig. 2B, a; dark and light red bars), defining an *A. thaliana* 'EC microbiome'. Of these 164, 97 were enriched in EC samples from both soil types (Fig. 2B, a; dark red bars), potentially representing a core EC microbiome. By contrast, 67 of these 164 were enriched in EC to a greater extent in one soil than the other (Fig. 2B, a; light red bars; Fig. 2B, b)). Importantly, 32 OTUs were depleted in EC samples (Fig. 2B, a; blue bars). Some OTUs exhibited rhizosphere enrichment; these significantly overlapped the EC-enriched OTUs ($P < 10^{-16}$, one-sided hypergeometric test) and also sometimes had a soil-type component (Fig. 2B, c and d). Only a few rhizosphere-specific enrichments were not also enriched in the EC (Supplementary Table 3). Hence, the *A. thaliana* EC microbiome is enriched for both a shared set of OTUs commonly assembled across two replicates from two diverse soils, and a set of OTUs that are assembled from each soil.

We assessed taxonomic distributions, first those of the 778 measurable OTUs in soil, rhizosphere and EC fractions, and then those of the 256 EC-enriched and 32 EC-depleted OTUs (Fig. 2A, Supplementary Fig. 7a and Supplementary Table 3). Measurable OTUs were distributed across seven dominant phyla (Fig. 2C and Supplementary Fig. 7c) and contained ~50–70% of the usable reads in all fractions (Supplementary Fig. 4c). Phyla distribution of the EC-enriched OTUs reflected that of the entire EC. Conversely, the phyla distribution of the EC-depleted OTUs typically resembled that of the rhizosphere fraction (Fig. 2C). The lower Shannon diversity of the EC fraction is consistent with enrichment for a subset of dominant phyla. Specifically, the EC microbiome was dominated by Actinobacteria, Proteobacteria and Firmicutes, and was depleted of Acidobacteria, Gemmatimonadetes and Verrucomicrobia, when soil types were considered either together or separately (Fig. 2C, Supplementary Figs 7c and 15 and Supplementary Table 5). Lower-order taxonomic analysis (Fig. 2D and Supplementary Fig. 7d) demonstrated that enrichment of a low-diversity Actinobacteria community in the EC was driven by a subset of families, predominantly Streptomycetaceae.

Other phyla, such as Proteobacteria, were represented by both EC enrichments and EC depletions at the family level (Fig. 2E and Supplementary Fig. 7e). Strikingly, two alphaproteobacterial families, Rhizobiaceae and Methylobacteriaceae, and two gammaproteobacterial families, Pseudomonadaceae and Moraxellaceae, dominated the EC population in their respective classes (Fig. 2F, α and γ, and Supplementary Fig. 7f, α and γ). Equally striking was the EC redistribution of particular alpha- and gammaproteobacterial families that were common in soil and rhizosphere (Fig. 2F and Supplementary Fig. 7f).

Specific OTUs, three from the family Streptomycetaceae and one from the order Sphingobacteriales, demonstrate the robustness of EC enrichments (Fig. 3a–d and Supplementary Fig. 11a–d). A few OTUs

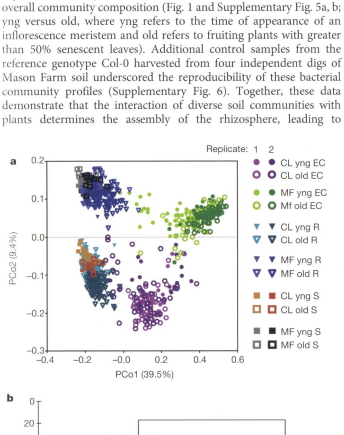


Figure 1 | Sample fraction and soil type drive the microbial composition of root-associated endophyte communities. a, Principal coordinate analysis of pairwise, normalized, weighted UniFrac distances between samples based on rarefaction to 1,000 reads in unthresholded, usable OTUs. CL, Clayton; MF, Mason Farm; R, rhizosphere; S, soil. **b**, Rarefied counts for the 25×5 thresholded, measurable OTUs from each of 24 soil, stage or fraction groups were $\log_2$-transformed (Methods) to make 24 representative samples (branch labels), and pairwise Bray–Curtis similarity was used to cluster these representatives hierarchically (group-average linkage).

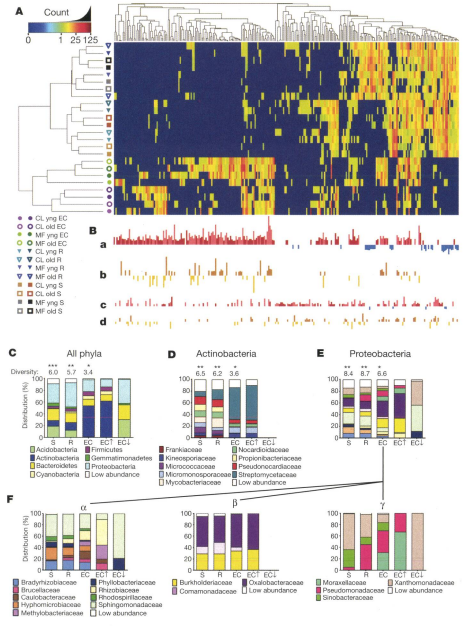


Figure 2 | OTUs that differentiate the EC and rhizosphere from soil. A, Heat map showing OTU counts from the rarefied OTU table (Supplementary Database 2a; $\log_2$-transformed) from each of the 256 rhizosphere- and EC-differentiating OTUs present across replicates. Samples and OTUs are clustered on their Bray–Curtis similarities (group-average linkage). The key relates colours to the untransformed read counts. Different hues of the same colour correspond to different replicates as in Fig. 1. **B**, The strength of GLMM predictions (best linear unbiased predictors) is represented by bar height. **a**, OTUs predicted as EC enriched (red, up) or EC depleted (blue, down). **b**, OTUs higher in the EC in Mason Farm soil than Clayton (brown, up) or higher in Clayton soil than Mason Farm (gold, down). OTUs in **a** that are not differentially affected by soil type are shown there in darker hues. **c**, OTUs predicted as rhizosphere enriched (as in **a**). **d**, OTUs higher in rhizosphere in one soil type (as in **b**). **C**, Histograms showing the distributions of phyla present in the 778 measurable OTUs in soil, rhizosphere and ECs compared with phyla present in the subset of EC OTUs enriched (EC↑) or depleted (EC↓) relative to soil. Shannon diversity (considering phyla as individuals) is given above each bar. A differential number of asterisks above the diversity values represents a significant difference ($P < 0.05$, weighted analysis of variance; Supplementary Methods and Supplementary Table 5). **D**, Distribution of families present among the OTUs from the phylum Actinobacteria. **E**, Distribution of families present among the OTUs from the phylum Proteobacteria. **F**, Distribution of families present among the OTUs of three classes of the phylum Proteobacteria: Alphaproteobacteria (α), Betaproteobacteria (β) and Gammaproteobacteria (γ). Statistical evidence for presence, enrichment in or depletion from EC is in Supplementary Table 6.

were either significantly enriched in rhizosphere but not in the EC (Fig. 3e, f, Supplementary Fig. 11e, f and Supplementary Table 3), or were associated with one of the two developmental stages (Fig. 3g, h, Supplementary Fig. 11g, h and Supplementary Table 3). Data in Fig. 2, Supplementary Fig. 7, Fig. 3, Supplementary Fig. 11 and Supplementary Table 3 demonstrate that entire taxa at various levels are enriched in or depleted from the EC microbiome. Additionally, rhizosphere taxa capable of colonizing the root vicinity are nonetheless prevented from colonizing the EC.

Several OTUs differentiated inbred *A. thaliana* accessions. Genotype-dependent enrichments and depletions were significant but weak (Supplementary Tables 5 and 3). To identify accession-dependent effects specific to a soil type or a developmental stage, we fitted a partial GLMM that modelled each genotype against bulk soil for each experiment or developmental stage group, and tested the model's predictions with a non-parametric Kruskal–Wallis test corrected for multiple testing (Supplementary Information). We considered only those significant accession-dependent effects that were present in the same direction in both biological replicates. We further required that these OTUs have a consistent prediction in the full GLMM, which narrowed the field to 12 OTUs (or 27 with frequency-normalized data; Supplementary Table 3). In Fig. 3, we display relative abundances of two such OTUs, one for each soil type, both Actinobacteria (Fig. 3i, j and Supplementary Fig. 11i, j). That these enrichments were detected by the full GLMM (which accounts for plate effects due to 454 sequencing), and were sequenced over several plates (Supplementary Fig. 14) supports a true genotype effect. Thus, a small subset of the EC microbiome is likely to be quantitatively influenced by host-genotype-dependent fine-tuning in specific soil environments. This could allow compensatory contributions of the EC microbiome and host genome variation to overall metagenome function.

Because the rhizoplane is stripped during preparation of EC fractions, we confirmed the presence of live bacteria on roots using catalysed reporter deposition and fluorescence *in situ* hybridization (CARD–FISH) to whole Col-0 root segments[18]. Eubacteria were common on unsonicated roots (Fig. 4a). Actinobacteria detected with probe HGC69a were visible on the surface of roots grown in Mason Farm soil, and co-localized with a subset of the eubacterial signals using double CARD–FISH (Fig. 4b), suggesting that their enrichment in EC fractions either comes from, or egresses through, the rhizoplane.

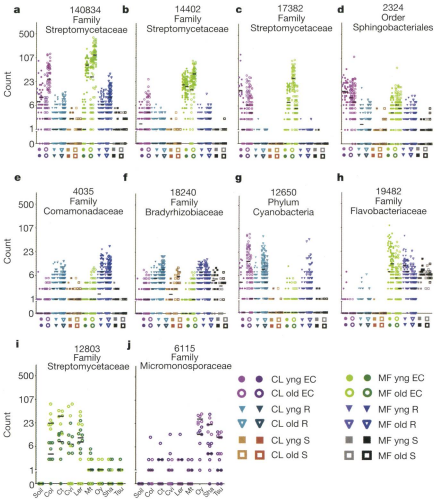


Figure 3 | Dot plots of notable OTUs. Counts for each OTU (number at top keyed to Supplementary Table 3) from the rarefied table were $\log_2$-transformed and the counts for each sample plotted as an individual symbol. The *y* axis is labelled with the actual (untransformed) counts. **a**–**h**, Each position on the *x* axis is labelled with a symbol to represent the sample group, and samples from that group are plotted in the column directly above. Biological replicates in the same column have different hues. The median of each replicate is shown with a horizontal black bar; some are invisible because they are at 0. **i**, **j**, Each *x*-axis position is labelled by *Arabidopsis* accession; samples from that accession are plotted above each label. Each OTU in the figure has model predictions in several categories (Supplementary Table 3).

Similarly, we confirmed the rare presence on the rhizoplane of Bradyrhizobiaceae (Supplementary Fig. 12c), a family with members defined by the GLMM as more abundant in Mason Farm rhizosphere than Mason Farm EC (Fig. 3f and Supplementary Fig. 11f). We enumerated the relative number of CARD–FISH signals on a set of filters made from equal amounts of material harvested in the same way as were the samples processed for pyrotag sequencing (Supplementary Fig. 12a, b). We confirmed that Actinobacteria were found in higher abundance, and that Bradyrhizobiaceae were present in lower abundances, in EC samples than in the bulk soil and rhizosphere samples. We also noted that emerging lateral roots were typically heavily colonized by a variety of bacteria (Supplementary Fig. 12d) consistent with previous observations[19]. These results are PCR-independent support for our sequencing methods.

We present a reduced-complexity, robust experimental platform with which to study root microbiota. Our data, and similar conclusions presented in a companion publication[20] using a similar platform, provide the deepest analysis available regarding the principles of root microbiome assembly for any plant species. Remarkably, our conclusions are very similar to those in ref. 20 and we identify phyla and family level enrichments in the EC fraction that largely overlap with those reported in ref. 20. We note three main differences between our study and that of ref. 20: different soils from a different continent, a different primer pair and a different portion of root harvested (top 3 cm in ref. 20; whole root here).

A subset of the soil bacterial population is typically enriched in rhizosphere samples[7]. Thus, a diverse bacterial community can surround the root surface and thrive there, recruited by biophysical and/or host-derived metabolic cues. We demonstrate that the *A. thaliana* microbiome undergoes dramatic loss of diversity as the spatial level of plant–microbe 'intimacy' further increases from the external rhizosphere to the intercellular EC. Both common and soil-type-specific OTUs are established inside roots grown in diverse soils. A small number of bacterial taxa, particularly the Actinobacteria family Streptomycetaceae, and several Proteobacteria families, are highly enriched in the EC. Actinobacteria are well known for production of antimicrobial secondary metabolites[9], and many proteobacterial families contain plant-growth-promoting members. Conversely,

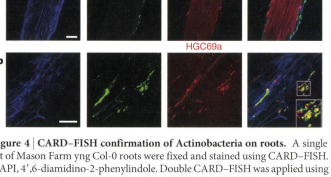


Figure 4 | CARD–FISH confirmation of Actinobacteria on roots. A single set of Mason Farm yng Col-0 roots were fixed and stained using CARD–FISH. DAPI, 4′,6-diamidino-2-phenylindole. Double CARD–FISH was applied using the EUB338 eubacterial probe (green) and either the NON338 probe (**a**), which is the nonsense negative control of EUB338, or the HGC69a Actinobacteria probe (**b**). Inset, twofold enlargement of boxed region. Scale bars, 50 μm.

several taxa (Acidobacteria, Verrucomicrobia and Gemmatimonadetes, and various proteobacterial families) that are common in soil and rhizosphere are depleted from the EC. This depletion suggests that these taxa are either actively excluded by the host immune system, outcompeted by more-successful EC colonizers or metabolically unable to colonize the EC niche. Our identification of a limited-diversity EC facilitates detailed characterization of the isolates comprising the core *A. thaliana* microbiome, which could facilitate the design of community-based plant probiotics.

Within the EC, we identified rare cases of quantitative variation in the enrichment of specific bacteria at two developmental stages or by different host genotypes, consistent with rare genotype-dependent associations noted in ref. 20. The former result suggests that the EC microbiome is robust to the source–sink differences across these two developmental stages, which may be related to the relatively high frequency of putative saprophytes defined in ref. 20. The latter result suggests that host genetic variation can drive either differential recruitment of beneficial microbes and/or differential exclusion. A limited-diversity EC microbiome with common features suggests similar host needs across *A. thaliana*, potentially extending to other plant taxa. These are probably fulfilled by contributions from a limited number of bacterial taxa across diverse soils. The identification of genotype-specific endophyte associations in particular soils may signal interactions that meet environment-specific host needs, balancing contributions of EC microbiome and host genome variation to overall metagenome function. These two generalities suggest that the *A. thaliana* root microbiome might assemble by core ecological principles similar to those shaping the mammalian microbiome, in which core phylum level enterotypes provide broad metabolic potential combined with modest levels of host-genotype-dependent associations that individualize the metagenome[21,22]. Isolation and characterization of the microbes that define host-genotype-dependent associations, and characterization beyond the 16S gene, should be particularly instructive in unravelling the molecular rules contributing to endophytic colonization and persistence.

METHODS SUMMARY

Custom methods of soil harvesting, seed sterilization and germination were developed to ensure no microbial carry over during transplantation into natural soils. Seedling growth and harvesting conditions were developed to maximize consistency. PCR primers were evaluated. The JGI multiplexed 454 sequencing pipeline was used to derive primary data, which was processed using standard methods and custom scripts. Analyses were performed on simplified data sets defined using a GLMM with statistical corrections. *In situ* methods were adapted to observe specific microbes defined by the phylotyping pipeline. All these steps are detailed in Supplementary Information.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 29 September 2011; accepted 15 May 2012.

1. Rodriguez, R. J. *et al.* Stress tolerance in plants via habitat-adapted symbiosis. *ISME J.* **2,** 404–416 (2008).
2. De Deyn, G. B., Cornelissen, J. H. C. & Bardgett, R. D. Plant functional traits and soil carbon sequestration in contrasting biomes. *Ecol. Lett.* **11,** 516–531 (2008).
3. van der Lelie, D. *et al.* Poplar and its bacterial endophytes: coexistence and harmony. *Crit. Rev. Plant Sci.* **28,** 346–358 (2009).
4. Jones, J. D. G. & Dangl, J. L. The plant immune system. *Nature* **444,** 323–329 (2006).
5. Dodds, P. N. & Rathjen, J. P. Plant immunity: towards an integrated view of plant-pathogen interactions. *Nature Rev. Genet.* **11,** 539–548 (2010).
6. Marschner, H., Römheld, V., Horst, W. J. & Martin, P. Root-induced changes in the rhizosphere: importance for the mineral nutrition of plants. *Z. Pflanz. Boden.* **149,** 441–456 (1986).
7. Dennis, P. G., Miller, A. J. & Hirsch, P. R. Are root exudates more important than other sources of rhizodeposits in structuring rhizosphere bacterial communities? *FEMS Microbiol. Ecol.* **72,** 313–327 (2010).
8. Mendes, R. *et al.* Deciphering the rhizosphere microbiome for disease-suppressive bacteria. *Science* **332,** 1097–1100 (2011).
9. Firáková, S., Šturdíková, M. & Múčková, M. Bioactive secondary metabolites produced by microorganisms associated with plants. *Biologia* **62,** 251–257 (2007).
10. Schulz, B. J. E., Boyle, C. J. C., Sieber, T. N., Schulz, B. & Boyle, C. in *Microbial Root Endophytes* Vol. 9, 1–13 (Springer, 2006).
11. Hallmann, J., Quadt-Hallmann, A., Mahaffee, W. F. & Kloepper, J. W. Bacterial endophytes in agricultural crops. *Can. J. Microbiol.* **43,** 895–914 (1997).
12. Redford, A. J., Bowers, R. M., Knight, R., Linhart, Y. & Fierer, N. The ecology of the phyllosphere: geographic and phylogenetic variability in the distribution of bacteria on tree leaves. *Environ. Microbiol.* **12,** 2885–2893 (2010).
13. Hardoim, P. R., van Overbeek, L. S. & Elsas, J. D. Properties of bacterial endophytes and their proposed role in plant growth. *Trends Microbiol.* **16,** 463–471 (2008).
14. Inceoglu, O., Salles, J. F., van Overbeek, L. & van Elsas, J. D. Effect of plant genotype and growth stage on the β-proteobacterial community associated with different potato cultivars in two fields. *Appl. Environ. Microbiol.* **76,** 3675–3684 (2010).
15. Inceoğlu, O., Al-Soud, W. A., Salles, J. F., Semenov, A. V. & van Elsas, J. D. Comparative analysis of bacterial communities in a potato field as determined by pyrosequencing. *PLoS ONE* **6,** e23321 (2011).
16. Benson, A. K. *et al.* Individuality in gut microbiota composition is a complex polygenic trait shaped by multiple environmental and host genetic factors. *Proc. Natl Acad. Sci. USA* **107,** 18933–18938 (2010).
17. Gottel, N. R. *et al.* Distinct microbial communities within the endosphere and rhizosphere of Populus deltoides roots across contrasting soil types. *Appl. Environ. Microbiol.* **77,** 5934–5944 (2011).
18. Eickhorst, T. & Tippkötter, R. Improved detection of soil microorganisms using fluorescence in situ hybridization (FISH) and catalyzed reporter deposition (CARD-FISH). *Soil Biol. Biochem.* **40,** 1883–1891 (2008).
19. Chi, F. *et al.* Ascending migration of endophytic rhizobia, from roots to leaves, inside rice plants and assessment of benefits to rice growth physiology. *Appl. Environ. Microbiol.* **71,** 7271–7278 (2005).
20. Bulgarelli, D. *et al.* Structure of and assembly cues for the *Arabidopsis* root-inhabiting bacterial microbiota. *Nature* doi:10.1038/nature11336 (this issue).
21. Arumugam, M. *et al.* Enterotypes of the human gut microbiome. *Nature* **473,** 174–180 (2011).
22. Spor, A., Koren, O. & Ley, R. Unravelling the effects of the environment and host genotype on the gut microbiome. *Nature Rev. Microbiol.* **9,** 279–290 (2011).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank A. Smithlund, M. Gonek, V. Madden, H. Schmidt, M. Rott and N. Zvenigorodsky for technical assistance. We thank A. Spor, J. Pfeiffer and J. Rawls for discussions and C. Herring for research field soil. This work was supported by US NSF grant IOS-0958245 (J.L.D.), the JGI Director's Discretionary Grand Challenge Program and the HHMI-GBMF Plant Science Program. J.L.D. is an HHMI-GBMF Plant Science Investigator. Work conducted by the US Department of Energy Joint Genome Institute is supported by the Office of Science of the US Department of Energy under contract no. DE-AC02-05CH11231. S.L.L. was supported by the National Institutes of Health, Minority Opportunities in Research division of the National Institute of General Medical Sciences grant K12GM000678.

Author Contributions D.S.L., S.L.L. and J.L.D. designed the project, D.S.L. and S.L.L. set up the experiments and organized construction of the sequencing libraries. D.S.L., S.L.L., S.H.P. and J.G. harvested samples and prepared DNA for sequencing. A.E., V.K., T.G.d.R., S.M., P.H. and S.G.T. worked together at the JGI to perform, optimize and quality-control the sequencing. S.H.P. applied the GLMM to the data. D.S.L., S.H.P., S.Y. and R.E. created and managed the data analysis pipeline. S.Y. oversaw the data deposition. D.S.L., S.L.L., S.H.P., S.Y., J.G. and J.L.D. analysed the data and created figures. S.L.L. performed the CARD–FISH microscopy in the laboratory of T.E., at the Max-Planck-Institute for Plant Breeding in Cologne, and at UNC. R.E.L. and P.H. advised on primer design and appropriate statistical methods. D.S.L., S.L.L., S.H.P. and J.L.D. wrote the manuscript with significant input from S.Y., R.E.L., P.H. and S.G.T.

Author Information 454 pyrosequencing data are deposited in the ENA Sequence Read Archive under study number ERP001384. Custom R scripts are available with documentation at http://labs.bio.unc.edu/dangl/resources/scripts. Lundberg_et_al_2012.htm and additional code is available on request. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to J.L.D. (dangl@email.unc.edu).

LETTER

doi:10.1038/nature11336

Revealing structure and assembly cues for *Arabidopsis* root-inhabiting bacterial microbiota

Davide Bulgarelli[1]*, Matthias Rott[1]*, Klaus Schlaeppi[1]*, Emiel Ver Loren van Themaat[1]*, Nahal Ahmadinejad[1]†, Federica Assenza[1], Philipp Rauf[1]†, Bruno Huettel[2], Richard Reinhardt[2], Elmon Schmelzer[3], Joerg Peplies[4], Frank Oliver Gloeckner[4,5], Rudolf Amann[5], Thilo Eickhorst[6] & Paul Schulze-Lefert[1]

The plant root defines the interface between a multicellular eukaryote and soil, one of the richest microbial ecosystems on Earth[1]. Notably, soil bacteria are able to multiply inside roots as benign endophytes and modulate plant growth and development[2], with implications ranging from enhanced crop productivity[3] to phytoremediation[4]. Endophytic colonization represents an apparent paradox of plant innate immunity because plant cells can detect an array of microbe-associated molecular patterns (also known as MAMPs) to initiate immune responses to terminate microbial multiplication[5]. Several studies attempted to describe the structure of bacterial root endophytes[6]; however, different sampling protocols and low-resolution profiling methods make it difficult to infer general principles. Here we describe methodology to characterize and compare soil- and root-inhabiting bacterial communities, which reveals not only a function for metabolically active plant cells but also for inert cell-wall features in the selection of soil bacteria for host colonization. We show that the roots of *Arabidopsis thaliana*, grown in different natural soils under controlled environmental conditions, are preferentially colonized by Proteobacteria, Bacteroidetes and Actinobacteria, and each bacterial phylum is represented by a dominating class or family. Soil type defines the composition of root-inhabiting bacterial communities and host genotype determines their ribotype profiles to a limited extent. The identification of soil-type-specific members within the root-inhabiting assemblies supports our conclusion that these represent soil-derived root endophytes. Surprisingly, plant cell-wall features of other tested plant species seem to provide a sufficient cue for the assembly of approximately 40% of the *Arabidopsis* bacterial root-inhabiting microbiota, with a bias for Betaproteobacteria. Thus, this root sub-community may not be *Arabidopsis*-specific but saprophytic bacteria that would naturally be found on any plant root or plant debris in the tested soils. By contrast, colonization of *Arabidopsis* roots by members of the Actinobacteria depends on other cues from metabolically active host cells.

We have grown *Arabidopsis* ecotypes Shakdara (Sha) and Landsberg *erecta* (Ler) in natural soils of contrasting geochemistry, designated Cologne (clay- and silt-rich) or Golm (sand- rich) soil, under controlled environmental conditions and at a defined planting density (Supplementary Fig. 1 and Supplementary Table 1). At early flowering stage we collected samples from three compartments: 'unplanted soil' (number of replicates: Cologne $n_C = 13$, Golm $n_G = 12$), 'rhizosphere' ($n_C = 15$, $n_G = 12$) and 'root' ($n_C = 18$, $n_G = 14$). The 'rhizosphere compartment' defines the soil particles firmly attached to roots collected by centrifugation of root washings (Supplementary Movie 1). The 'root compartment' is defined as root tissue depleted of soil particles and epiphytic bacteria by sequential washing and sonication treatments and is therefore enriched for root-inhabiting bacteria (Supplementary Fig. 2). We used pyrosequencing of an approximately 400 base pairs PCR amplicon of the bacterial 16S ribosomal RNA gene and analysed the variable gene segments V5–V6.

To examine the taxonomic structure of the bacterial communities we performed a supervised taxonomy classification of all high quality reads using the SILVA[7] database. This classification identified a total of 43 bacterial phyla and divisions and revealed an anomalous Chloroflexi abundance in all samples (Fig. 1a). PCR-independent catalysed reporter deposition-fluorescence *in situ* hybridization (CARD-FISH) analysis on soil samples (Supplementary Fig. 3) and comparative PCR primer analysis indicated this is due to a PCR primer bias (Supplementary Information and Supplementary Fig. 4). After removal of reads assigned to Chloroflexi we identified Proteobacteria, Actinobacteria and Bacteroidetes as dominating phyla in root bacterial communities and significantly enriched compared to soil and

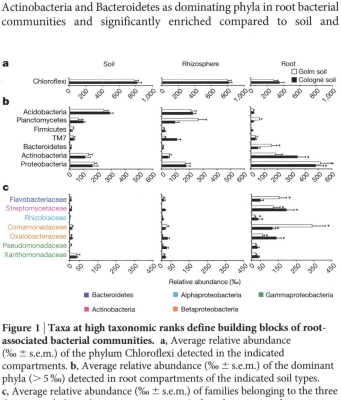


Figure 1 | Taxa at high taxonomic ranks define building blocks of root-associated bacterial communities. a, Average relative abundance (‰ ± s.e.m.) of the phylum Chloroflexi detected in the indicated compartments. **b**, Average relative abundance (‰ ± s.e.m.) of the dominant phyla (> 5‰) detected in root compartments of the indicated soil types. **c**, Average relative abundance (‰ ± s.e.m.) of families belonging to the three dominant phyla in the root compartment. In **b** and **c** average relative abundances are calculated after removal of reads assigned to Chloroflexi. Asterisks indicate significant enrichment (Benjamini–Hochberg false-discovery-rate (FDR) adjusted P value < 0.05) in the root compartment compared to soil and rhizosphere compartments.

[1]Department of Plant Microbe Interactions, Max Planck Institute for Plant Breeding Research, 50829 Cologne, Germany. [2]Max Planck Genome Centre, Max Planck Institute for Plant Breeding Research, 50829 Cologne, Germany. [3]Central Microscopy, Max Planck Institute for Plant Breeding Research, 50829 Cologne, Germany. [4]Ribocon GmbH, 28359 Bremen, Germany. [5]Department of Molecular Ecology, Max Planck Institute for Marine Microbiology, 28359 Bremen, Germany. [6]Soil Science, Faculty of Biology and Chemistry, University of Bremen, 28359 Bremen, Germany. †Present addresses: INRES - Crop Bioinformatics, University of Bonn, 53115 Bonn, Germany (N.A.); Department of Developmental Immunology, Max Planck Institute of Immunobiology and Epigenetics, 79108 Freiburg, Germany (P.R.).

*These authors contributed equally to this work.

rhizosphere (Fig. 1b). Within the root-inhabiting Proteobacteria we noted an over-representation of Betaproteobacteria families compared to Alphaproteobacteria and Gammaproteobacteria (Fig. 1c). Likewise, in each of the other three root-inhabiting phyla we noted a dominating family: Flavobacteriaceae for the phylum Bacteroidetes and Streptomycetaceae for the phylum Actinobacteria (Fig. 1c). The Streptomycetaceae and the families belonging to the Betaproteobacteria are significantly enriched in the root compared to unplanted soil and rhizosphere compartments in both soils tested (Fig. 1c and Supplementary Data 1). Together this indicates that building blocks of *Arabidopsis* root-inhabiting bacterial communities are detected at the family level.

We noted that unplanted soil and rhizosphere contain a disproportionate number of reads that cannot be unambiguously classified at the order level using two different databases (Supplementary Fig. 5 and Supplementary Information), indicative of an insufficient database representation of the biodiversity of soil-borne bacteria[4]. To overcome this limitation, we clustered the 16S rRNA gene sequences of all compartments and defined operational taxonomic units (OTUs) of bacteria at ≥ 97% sequence identity. Bacterial diversity measured as OTU richness was estimated in the three compartments by rarefaction analysis and revealed the greatest number of OTUs in unplanted soil (~2,000), followed by a reduced richness in rhizosphere and root compartments (each ~1,000) (Supplementary Fig. 6). Technical replicates of the same DNA sample defined a minimum threshold of 5‰ relative abundance for reproducible quantification of individual OTUs (Supplementary Fig. 7). For subsequent analysis we also depleted OTUs assigned to the phylum Chloroflexi and consequently nine samples with less than 1,000 reads were excluded (Supplementary Information).

To compare the composition of the identified community members we generated a hierarchical cluster based on Bray–Curtis distance (Supplementary Fig. 8). Consistent with studies using other plant species, this showed a striking effect of the host on associated bacterial microbiota[9–11]. The root compartment renders the root-inhabiting microbiota significantly dissimilar from the communities retrieved from rhizosphere and unplanted soil compartments (Supplementary Fig. 8 and Supplementary Table 2). Closer examination also showed a marked soil-type-dependent effect on both unplanted soil and rhizosphere communities, indicating a different natural bacterial start inoculum in the tested soils. A differentiation between unplanted soil and the rhizosphere microbiota across seasonal soil batches was not evident from the cluster dendrogram (Supplementary Fig. 8). However, a distinct bacterial community in the rhizosphere compartment is detectable when looking at samples obtained from the same soil batch (Supplementary Fig. 9). This differentiation is reflected by a significant rhizosphere effect according to PERMANOVA analysis of the Bray–Curtis distance matrix (Supplementary Table 2), indicating that a rhizosphere effect is obscured by soil batch-to-batch variation.

To identify bacteria responsible for the observed community differentiation (Supplementary Fig. 8), we used a linear model analysis to determine indicator OTUs for each tested compartment (Supplementary Data 1). Consistent with the Bray–Curtis cluster analysis, we found a subset of OTUs significantly enriched in roots, whereas unplanted soil and rhizosphere compartments share a large proportion of OTUs (Fig. 2a).

To obtain insights in potential plant-derived assembly cues for the root-inhabiting microbiota, we incubated untreated wooden splinters representing metabolically inactive lignocellulosic matrices for bacterial colonization (Supplementary Figs 1 and 10) as additional compartment in both tested soils. We determined the associated bacterial communities of the softwood birch (*Betula*, $n_C = 8$; $n_G = 11$) and the hardwood beech (*Fagus*, $n_C = 4$; $n_G = 4$). Remarkably, approximately 40% of the *Arabidopsis* root-enriched OTUs in Cologne soil are equally or even more abundant in these wooden compartments (Fig. 2b, c and Supplementary Data 1). The identification of this shared subcommunity (designated lOTUs for lignocellulosic matrix-associated

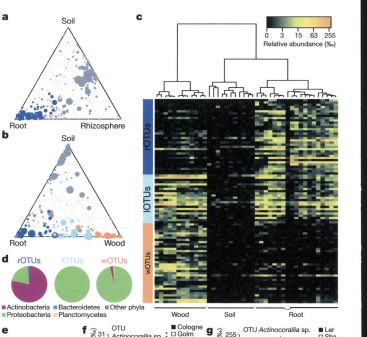


Figure 2 | *Arabidopsis* assembles a distinctive root-inhabiting bacterial microbiota. a–g, Compartment specificity and relative abundance of OTUs (> 5‰) determined in samples from Cologne soil (**a–d**) and in comparison to Golm soil (**e–g**). **a**, Ternary plot of all OTUs. Each circle represents one OTU. The size of each circle represents its relative abundance (weighted average). The position of each circle is determined by the contribution of the indicated compartments to the total relative abundance. The dotted grid and numbers inside the plot indicate 20% increments of contribution from each compartment (see Supplementary Methods). Dark blue circles mark OTUs significantly enriched in the root compartment (FDR < 0.05). **b**, Ternary plot similar to **a** including the wood compartment. OTUs significantly enriched in root (dark blue, rOTUs), wood-enriched community members (orange, wOTUs) and OTUs shared by root and wood compartments (light blue, lOTUs) (FDR < 0.05). **c**, Heat map of the relative abundance of root- and wood-enriched OTUs. Vertical columns represent samples, horizontal rows depict OTUs. Clustering of samples (top) is based on OTUs co-occurrence. Colour code on the left indicates OTU compartment specificity as defined in **b**. **d**, Taxonomic composition of rOTUs, lOTUs and wOTUs subcommunities. The size of each segment in the chart is proportional to the cumulative relative abundance of OTUs assigned to the indicated taxa. **e**, Numbers of lOTUs and rOTUs in the indicated soils (FDR < 0.05). **f**, relative abundance of OTU *Actinocorallia* sp. in the indicated compartments (mean ± s.e.m.). Asterisks indicate significant enrichment in the root compartment over all other indicated compartments (FDR < 0.05). **g**, Differential relative abundance of OTU *Actinocorallia sp.* in the root compartment of the indicated *Arabidopsis* ecotypes (mean ± s.e.m.). Asterisks indicate significant differences (FDR < 0.05).

OTUs) indicates that plant cell-wall features serve as sufficient colonization cue. A second sub-community, specifically enriched in roots (designated rOTUs), seems to depend on other or additional cues from metabolically active host cells (Fig. 2b, c and Supplementary Data 1). A third sub-community is specifically enriched in the wooden matrices (designated wOTUs; Fig. 2b, c and Supplementary Data 1). Under-representation of wOTUs in roots might reflect a selective inhibitory plant activity against colonization by these bacteria. The three sub-communities identified through a comparison of root and wooden compartments were also seen in analogous experiments in Golm soil (Supplementary Fig. 11 and Supplementary Data 1). Distinctive root- and wood-derived OTU profiles and subtle differences between soil- and rhizosphere-derived profiles were further

supported by the classifiability of the compartments (Supplementary Fig. 12 and Supplementary Table 3), non-parametric tests (Supplementary Fig. 13) and were robust against the primer bias (Supplementary Fig. 14). Notably, a different taxonomic structure defines the three sub-communities. Whereas Proteobacteria represent the vast majority of wOTUs and lOTUs community members, Actinobacteria were largely underrepresented in these communities compared to the rOTUs community (Fig. 2d and Supplementary Fig. 11). Hence, Actinobacteria are specifically enriched in roots. To examine a potential role of soil type on the root-inhabiting bacterial assemblage, we compared the bacterial profiles obtained in Cologne and Golm soils (Fig. 2e).

Within the rOTUs community, 9 OTUs were enriched in roots of both tested soil types, whereas 21 and 13 OTUs were significantly enriched in roots derived from Cologne or Golm soils, respectively (Fig. 2e). Notably, closer examination validates for 15 of the root-inhabiting OTUs (lOTUs plus rOTUs) a significant differential enrichment between the soil types (Supplementary Fig. 15 and Supplementary Data 1). This identifies soil type as an important environmental variable influencing the quantitative and qualitative composition of the *Arabidopsis* root-inhabiting microbiota. This also points to soil-type-dependent root colonization processes and/or reflects different bacterial start inocula.

A root-enriched OTU in Cologne and Golm soil, classified as *Actinocorallia* sp., belonging to the Actinobacteria phylum (Fig. 2f), showed a differential accumulation between the two *Arabidopsis* ecotypes tested, Sha and Ler (Fig. 2g). An approximately tenfold difference in relative abundance was observed in independent experiments using Cologne soil collected during fall or spring. A similar differential accumulation trend was found between these ecotypes grown in Golm soil although statistically not significant. Using a similar experimental platform[12], a small number of OTUs (12) of the root-inhabiting bacterial assemblage whose accumulation is quantitatively influenced by the host genotype in specific soil environments was identified among eight *Arabidopsis* ecotypes tested[14]. End-point PCR analysis using PCR primers designed on the basis of OTU *Actinocorallia* sp.-representative sequence that are specific for Actinomycetales (Supplementary Table 4) independently validated their presence in soil-grown roots of Sha and Ler (Supplementary Fig. 16). PCR amplicons were also detectable in unplanted soil samples but not in surface-sterilized and crushed seeds of either accession (Supplementary Fig. 16), indicating that the root-inhabiting *Actinocorallia* are recruited from soil. These findings indicate that natural genetic variation in *A. thaliana* exerts a quantitative control on root-inhabiting bacteria and that its phenotypic variation is influenced by the environmental component soil type.

The *Arabidopsis* rhizoplane, the interface between host tissue and rhizosphere soil (Supplementary Fig. 2), was largely eliminated by sonication during sample preparation for the pyrosequencing-based 16S rRNA gene survey. Therefore, we used scanning electron microscopy (SEM) and adopted a modified fluorescence *in situ* hybridization (FISH) protocol (CARD-FISH; Supplementary Information) to characterize and visualize bacteria attached to the rhizoplane. SEM analysis revealed morphologically diverse bacteria-like structures (Fig. 3 a–e) and CARD-FISH identified a high density of bacteria using the probe EUB338 that detects the majority of known Eubacteria corresponding to their rRNA content, and therefore represents a proxy of their metabolic state (Fig. 3f and Supplementary Table 5). Particulate CARD-FISH signals were not found using the reverse complement probe or upon hybridization with roots of axenically grown plants (Fig. 3g and Supplementary Fig. 17). All three taxa that dominate the root-inhabiting community based on pyrosequencing of PCR-amplified DNA were also detected on the rhizoplane by CARD-FISH (Betaproteobacteria, Bacteroidetes and Actinobacteria; Fig. 3h–j). Each CARD-FISH probe detected a distinct colonization pattern for the respective phylum. The same phylum-specific CARD-FISH probes detected very few signals in unplanted soil and rhizosphere samples

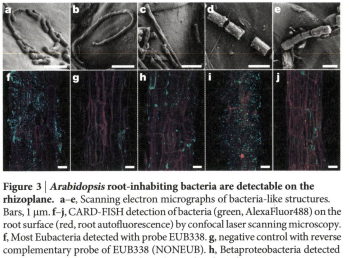


Figure 3 | *Arabidopsis* root-inhabiting bacteria are detectable on the rhizoplane. a–e, Scanning electron micrographs of bacteria-like structures. Bars, 1 μm. **f–j**, CARD-FISH detection of bacteria (green, AlexaFluor488) on the root surface (red, root autofluorescence) by confocal laser scanning microscopy. **f**, Most Eubacteria detected with probe EUB338. **g**, negative control with reverse complementary probe of EUB338 (NONEUB). **h**, Betaproteobacteria detected with probe BET42a. **i**, Bacteroidetes detected with probe CF319a. **j**, Actinobacteria detected with probe HGC69a. Bars, 20 μm.

compared to the EUB338 probe (Supplementary Fig. 3). Together with undetectable 16S rRNA gene amplicons from surface-sterilized and crushed seeds or roots from axenically grown plants (Supplementary Fig. 16), this indicates that these rhizoplane-attached bacteria are derived from the start inoculum of soil bacteria.

Are the results from our controlled environment experiments relevant under natural conditions? We collected roots of *A. thaliana* plants naturally grown at a site close to Cologne and determined bacterial profiles in the root, rhizosphere and corresponding soil compartments (Supplementary Information and Supplementary Data 1). These samples differ in many aspects from the controlled environment samples including vegetation period, soil geochemistry (Supplementary Table 1), climatic conditions, inter-species competition, uncontrolled biotic/abiotic stresses and an unknown *A. thaliana* genotype different from Ler and Sha accessions (Supplementary Fig. 18). Despite the many differences to the controlled environment experiments, the distinctiveness of the root-inhabiting microbiota from those of rhizosphere and unplanted soil compartments is retained in naturally grown *Arabidopsis* (Fig. 4a). Bray–Curtis analysis of the combined data from greenhouse and naturally grown plants confirms that the compartments are the major determinants of community structure. Notably, a large proportion of OTUs retrieved under greenhouse conditions were also detected in each of the three tested compartments of naturally grown *Arabidopsis* plants (Supplementary Fig. 19). This includes OTUs belonging to the root-enriched rOTUs and lOTUs sub-communities of the greenhouse experiments (compare Fig. 2e and Fig. 4b). Because these OTUs were preferentially detected in the root compartment of the natural site (Fig. 4c), the selectivity of *Arabidopsis* roots for the rOTUs seems robust in a natural and controlled environment. For example, the enrichment gradient found between the three tested compartments for the root-specific OTU *Actinocorallia sp.* in the natural specimens is similar to the one under controlled environmental conditions (Fig. 4d, compare to Fig. 2f).

Reproducible measurements of bacterial microbiota under controlled environmental conditions revealed a function for metabolically active plant cells and cell-wall features in the selection of soil bacteria for host colonization. Root and wooden compartments are preferentially colonized by Betaproteobacteria and Bacteroidetes (Fig. 2d and Supplementary Fig. 11). Members of these two phyla are characterized as copiotrophic soil bacteria, that is, they compete successfully only when organic resources are abundant[13]. Thus, colonization of *Arabidopsis* roots by the lOTUs sub-community could reflect their ability to proliferate in the presence of polysaccharide polymers and might contribute to the decomposition of organic matter after plant death. Our study predicts that the lOTUs sub-community is not

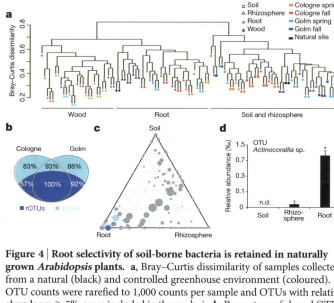


Figure 4 | Root selectivity of soil-borne bacteria is retained in naturally grown *Arabidopsis* plants. **a**, Bray–Curtis dissimilarity of samples collected from a natural (black) and controlled greenhouse environment (coloured). OTU counts were rarefied to 1,000 counts per sample and OTUs with relative abundance $> 5\%$ were included in the analysis. **b**, Percentage of shared OTUs between root-associated sub-communities found at controlled environmental conditions in Cologne or Golm soils and naturally grown *Arabidopsis* plants (comparison based on numbers shown in Fig. 2e). **c**, Ternary plot of the relative abundance and proportional contribution of OTUs ($> 5\%$) in the indicated compartments displaying lOTUs (light blue) and rOTUs (dark blue) identified in greenhouse experiments. **d**, Relative abundance of OTU *Actinocorallia sp.* (mean $\pm$ s.e.m.; n.d., not detected) in the indicated compartments. Asterisks indicate a significant difference (FDR < 0.05) between root and rhizosphere.

Arabidopsis-specific but may be saprophytic bacteria that would naturally be found on any plant root or plant debris in the tested soils. In contrast, the selective enrichment of Actinobacteria in the rOTUs community (Fig. 2d, Supplementary Fig. 11 and Supplementary Data 1) indicates that cues from metabolically active host cells are needed for the assembly of the rOTUs sub-community. Together with the observation that a subset of soil bacteria is enriched in the wooden compartment compared to roots, our results point to an active host process mediating attractant and repellent activities. Members of the rOTUs sub-community may provide probiotic functions for the plant. For example, Actinobacteria are known to produce a vast diversity of antimicrobial compounds[14].

The composition of the putative root endophyte microbiota is influenced by the soil type (Fig. 2e and Supplementary Fig. 15), which could reflect different natural start inocula in the three tested soils and supports our conclusion that at least a subset of these communities represents *Arabidopsis* root endophytes originating from the microbiota reservoir present in natural soil. The host genotype was found to have a limited effect on the root endophyte profile (Fig. 2g), which is reminiscent of the mouse gut microbiota for which the host genotype quantitatively contributes to its structure[15]. Strikingly similar findings are reported in ref. 12 using two additional soil types, eight *Arabidopsis* accessions and a similar fractionation protocol, but different PCR primers and different computational pipelines, thereby supporting the generality of our conclusions. Thus, our work provides a foundation for future molecular studies elucidating how *Arabidopsis* roots control and tolerate colonization by a specific endophyte community despite an elaborate innate immune system, including receptors for conserved bacterial structures such as flagellin[16].

METHODS SUMMARY

A detailed description of the natural soils and all methods used in this study can be found in the Supplementary Information. *Arabidopsis thaliana* ecotypes Shakdara and Landsberg *erecta* were grown in pots filled with natural Cologne or Golm soil under long-day conditions (16-h photoperiod) at a defined planting density and roots were harvested at early flowering stage. Wooden *Fagus* and *Betula* splinters were inserted into soil to a depth of approximately 4 cm and as well as unplanted soil pots subjected to the same conditions as pots with living plants. The uppermost 3 cm of roots and soil-incubated wood were harvested and adhering soil was washed off and defined as 'rhizosphere compartment'. After washing, roots and wooden splinters were sonicated to remove the bacterial surface biofilm and to enrich for endophytic bacteria ('root'/'wood' compartments). DNA extraction of all compartments was performed using the MP Bio Fast DNA for Soil Kit. Barcoded bacterial 16S rRNA gene PCR amplicons were generated using a modified version of previously described primers[17,18] in combination with a touchdown PCR programme (Supplementary Table 6) to minimize host rRNA gene amplification. Amplicons were gel-purified (Qiagen), pooled and sequenced on a 454 Titanium platform (Roche). We performed a classification of the 454 reads using the SILVA[7] database. For OTU-based analysis we used PyroTagger[19] to screen for high-quality sequences and clusters of 97% sequence identity. Statistical analyses were performed using a series of packages and scripts developed in R. Significant differences of OTU- or SILVA-based taxonomy counts between samples from two compartments or in interaction terms were obtained using moderated *t*-tests on log-transformed relative abundance and corrected for multiple hypothesis testing. CARD-FISH experiments were performed as previously described with minor modifications[20,21]. SEM micrographs were recorded as described previously[22].

Received 28 September 2011; accepted 18 June 2012.

1. Tringe, S. G. *et al.* Comparative metagenomics of microbial communities. *Science* **308,** 554–557 (2005).
2. Hardoim, P. R., van Overbeek, L. S. & Elsas, J. D. Properties of bacterial endophytes and their proposed role in plant growth. *Trends Microbiol.* **16,** 463–471 (2008).
3. Mei, C. & Flinn, B. S. The use of beneficial microbial endophytes for plant biomass and stress tolerance improvement. *Recent Pat. Biotechnol.* **4,** 81–95 (2010).
4. Weyens, N., van der Lelie, D., Taghavi, S. & Vangronsveld, J. Phytoremediation: plant–endophyte partnerships take the challenge. *Curr. Opin. Biotechnol.* **20,** 248–254 (2009).
5. Rosenblueth, M. & Martínez-Romero, E. Bacterial endophytes and their interactions with hosts. *Mol. Plant Microbe Interact.* **19,** 827–837 (2006).
6. Reinhold-Hurek, B. & Hurek, T. Living inside plants: bacterial endophytes. *Curr. Opin. Plant Biol.* **14,** 435–443 (2011).
7. Pruesse, E. *et al.* SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res.* **35,** 7188–7196 (2007).
8. Gans, J., Wolinsky, M. & Dunbar, J. Computational improvements reveal great bacterial diversity and high metal toxicity in soil. *Science* **309,** 1387–1390 (2005).
9. Berg, G. & Smalla, K. Plant species and soil type cooperatively shape the structure and function of microbial communities in the rhizosphere. *FEMS Microbiol. Ecol.* **68,** 1–13 (2009).
10. İnceoğlu, Ö., Al-Soud, W. A., Salles, J. F., Semenov, A. V. & van Elsas, J. D. Comparative analysis of bacterial communities in a potato field as determined by pyrosequencing. *PLoS ONE* **6,** e23321 (2011).
11. Hardoim, P. R. *et al.* Rice root-associated bacteria: insights into community structures across 10 cultivars. *FEMS Microbiol. Ecol.* **77,** 154–164 (2011).
12. Lundberg, D. S. *et al.* Defining the core *Arabidopsis thaliana* root microbiome. *Nature* http://dx.doi.org/10.1038/nature11237 (this issue).
13. Fierer, N., Bradford, M. A. & Jackson, R. B. Toward an ecological classification of soil bacteria. *Ecology* **88,** 1354–1364 (2007).
14. Basilio, A. *et al.* Patterns of antimicrobial activities from soil actinomycetes isolated under different conditions of pH and salinity. *J. Appl. Microbiol.* **95,** 814–823 (2003).
15. Benson, A. K. *et al.* Individuality in gut microbiota composition is a complex polygenic trait shaped by multiple environmental and host genetic factors. *Proc. Natl Acad. Sci. USA* **107,** 18933–18938 (2010).
16. Boller, T. & Felix, G. A renaissance of elicitors: perception of microbe-associated molecular patterns and danger signals by pattern-recognition receptors. *Annu. Rev. Plant Biol.* **60,** 379–406 (2009).
17. Chelius, M. K. & Triplett, E. W. The diversity of archaea and bacteria in association with the roots of *Zea mays* L. *Microb. Ecol.* **41,** 252–263 (2001).
18. Buchholz-Cleven, B. E. E., Rattunde, B. & Straub, K. L. Screening for genetic diversity of isolates of anaerobic Fe(II)-oxidizing bacteria using DGGE and whole-cell hybridization. *Syst. Appl. Microbiol.* **20,** 301–309 (1997).
19. Kunin, V. & Hugenholtz, P. PyroTagger: a fast, accurate pipeline for analysis of rRNA amplicon pyrosequence data. *Open J. Article–1* (2010).
20. Eickhorst, T. & Tippkötter, R. Improved detection of soil microorganisms using fluorescence *in situ* hybridization (FISH) and catalyzed reporter deposition (CARD-FISH). *Soil Biol. Biochem.* **40,** 1883–1891 (2008).
21. Pernthaler, A., Pernthaler, J. & Amann, R. Fluorescence *in situ* hybridization and catalyzed reporter deposition for the identification of marine bacteria. *Appl. Environ. Microbiol.* **68,** 3094–3101 (2002).
22. Timmusk, S., Grantcharova, N. & Wagner, E. G. H. *Paenibacillus polymyxa* invades plant roots and forms biofilms. *Appl. Environ. Microbiol.* **71,** 7292–7300 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank R. Franzen and S. Schumacher for technical assistance, S. Wulfert for providing *Arabidopsis* liquid cultures, H. Schmidt for sharing the CARD-FISH protocol and performing the cell counts. We thank S. Klages and K. Stüber for support with bioinformatic pipelines. We are grateful to R. Panstruga and P. Bakker for comments on the manuscript. Golm soil was shipped by J. Schwachtje and J. van Dongen. This work was supported by funds from the Max Planck Society to P.S.-L. (M.IF. A. ZUCH8048). K.S. is supported by the Swiss National Science Foundation (PBFRP3-133544).

Author Contributions D.B., M.R., K.S., F.A., and E.V.L.v.T performed the experiments and analysed the data. M.R. and T.E. performed CARD-FISH experiments. E.S. generated SEM micrographs. E.V.L.v.T. coordinated computational analyses. N.A. performed the RDP classification. J.P. performed the SILVA classification. F.O.G. provided access to SILVA classification and R.A. advice on CARD-FISH, SILVA analysis. B.H. and R.R. performed pyrosequencing of amplicon libraries. D.B., M.R., K.S., P.R. and P.S.-L. designed the study. D.B., M.R., K.S., E.V.L.v.T and P.S.-L. wrote the manuscript. D.B., M.R., K.S. and E. V.L.v.T. contributed equally to this work. All authors discussed the results and commented on the manuscript.

Author Information Pyrosequencing reads have been deposited in the NCBI Sequence Read Archive (SRA) database (SRA043581). The R scripts used for computational analyses are available via http://www.mpipz.mpg.de/162701/R_scripts. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to P.S.-L. (schlef@mpipz.mpg.de).

LETTER

doi:10.1038/nature11283

A mutation in *APP* protects against Alzheimer's disease and age-related cognitive decline

Thorlakur Jonsson[1], Jasvinder K. Atwal[2], Stacy Steinberg[1], Jon Snaedal[3], Palmi V. Jonsson[3,8], Sigurbjorn Bjornsson[3], Hreinn Stefansson[1], Patrick Sulem[1], Daniel Gudbjartsson[1], Janice Maloney[2], Kwame Hoyte[2], Amy Gustafson[2], Yichin Liu[2], Yanmei Lu[2], Tushar Bhangale[2], Robert R. Graham[2], Johanna Huttenlocher[1,4], Gyda Bjornsdottir[1], Ole A. Andreassen[5], Erik G. Jönsson[6], Aarno Palotie[7], Timothy W. Behrens[2], Olafur T. Magnusson[1], Augustine Kong[1], Unnur Thorsteinsdottir[1,8], Ryan J. Watts[2] & Kari Stefansson[1,8]

The prevalence of dementia in the Western world in people over the age of 60 has been estimated to be greater than 5%, about two-thirds of which are due to Alzheimer's disease[1–4]. The age-specific prevalence of Alzheimer's disease nearly doubles every 5 years after age 65, leading to a prevalence of greater than 25% in those over the age of 90 (ref. 3). Here, to search for low-frequency variants in the amyloid-β precursor protein (*APP*) gene with a significant effect on the risk of Alzheimer's disease, we studied coding variants in *APP* in a set of whole-genome sequence data from 1,795 Icelanders. We found a coding mutation (A673T) in the *APP* gene that protects against Alzheimer's disease and cognitive decline in the elderly without Alzheimer's disease. This substitution is adjacent to the aspartyl protease β-site in APP, and results in an approximately 40% reduction in the formation of amyloidogenic peptides *in vitro*. The strong protective effect of the A673T substitution against Alzheimer's disease provides proof of principle for the hypothesis that reducing the β-cleavage of APP may protect against the disease. Furthermore, as the A673T allele also protects against cognitive decline in the elderly without Alzheimer's disease, the two may be mediated through the same or similar mechanisms.

Amyloid plaques are a central pathological feature of Alzheimer's disease and largely consist of amyloid-β peptides[5,6]. Amyloid-β is formed through sequential proteolytic processing of APP, catalysed by the β- and γ-secretases[7]. The aspartyl protease β-site APP cleaving enzyme 1 (BACE1), originally identified over a decade ago[8–11], cleaves APP predominantly at a unique site, whereas the γ-secretase complex cleaves the resulting carboxy-terminal fragment at several sites, with preference for positions 40 and 42, leading to formation of amyloid-β_{1-40} ($A\beta_{1-40}$) and $A\beta_{1-42}$ peptides[7]. Alternative processing of APP at the α-site prevents the formation of amyloid-β, as the α-site is located within amyloid-β.

Over 30 coding mutations in the *APP* gene have been found. About 25 of these are pathogenic, in most cases resulting in autosomal dominant Alzheimer's disease with an early onset[12,13]. Substitutions at or near the β- and γ-proteolytic sites appear to result in overproduction of either total amyloid-β or a shift in the $A\beta_{1-40}$:$A\beta_{1-42}$ ratio towards formation of the more toxic $A\beta_{1-42}$ peptide, whereas substitutions within the amyloid-β peptide are believed to result in formation of amyloid-β with increased propensity for aggregation[14].

Until now, mutations in *APP* have not been implicated in the common, late-onset form of Alzheimer's disease, with the exception of the rare variant, N660Y, which was recently identified in one case from a late-onset Alzheimer's disease family[15]. To search for low-frequency variants in the *APP* gene with a significant effect on the risk of Alzheimer's disease, we tabulated coding variants in *APP* in a set of whole-genome sequence data from 1,795 Icelanders. Variants present in more than one individual were subsequently imputed into 71,743 chip-typed Icelanders using long-range phasing information[16–18], followed by propagation of genotypes and generation of *in silico* genotypes for 296,496 close relatives of chip-typed individuals who had not been genotyped[19].

We then investigated the association of the variants in *APP* with Alzheimer's disease (Supplementary Table 1). The control group included individuals who had lived to at least age 85 without a diagnosis of Alzheimer's disease. The most significant association was found with rs63750847. The A allele of this single nucleotide polymorphism (SNP) (rs63750847-A) results in an alanine to threonine substitution at position 673 in APP (A673T), and was found to be significantly more common in the elderly control group than in the Alzheimer's disease group (0.62% versus 0.13%; odds ratio (OR) = 5.29; P value = 4.78×10^{-7}; Table 1), and is therefore protective against Alzheimer's disease. To confirm these results, we performed Sanger sequencing of rs63750847 in 451 predicted carriers of rs63750847-A, including two predicted homozygotes. All the predicted carriers were found to have the correct copy number of rs63750847-A, confirming the results obtained with imputation. We also confirmed results by genotyping rs63750847 in 3,661 individuals (cases and controls), and found one mismatch (0.027%; Supplementary Information). Previously, the rs63750847 variant had been reported in a single individual without a history of Alzheimer's disease[20] and in one affected member of a family with late-onset Alzheimer's disease, but was deemed to be probably non-pathogenic[15]. We found the variant in 3 out of 712

Table 1 | APP A673T protects against Alzheimer's disease

Analysis	1/OR	OR	P value	Controls		
				Frequency (%)	N_{chip}	$N_{in\ silico}$
AD	-	-	-	0.13	2,199	849
AD versus population controls	4.24	0.236	4.19×10^{-5}	0.45	57,174	22,074
AD versus population controls aged 85 or greater	5.29	0.189	4.78×10^{-7}	0.62	7,653	1,350
AD versus cognitively intact controls at age 85	7.52	0.133	6.92×10^{-6}	0.79	827	407

The table shows association results, comparing patients with Alzheimer's disease (AD) to three different control groups (top line gives numbers for patients with Alzheimer's disease only). N_{chip}, number of individuals with chip-based genotype information; $N_{in\ silico}$, number of individuals with genealogy-based genotype information.

[1]deCODE genetics, Sturlugata 8, 101 Reykjavik, Iceland. [2]Genentech, 1 DNA Way, South San Francisco, California 94080, USA. [3]Landspitali University Hospital, Department of Geriatrics, 101 Reykjavik, Iceland. [4]Department of Medical Genetics, Institute for Human Genetics, 72026 Tübingen, Germany. [5]Department of Psychiatry, Ullevål University Hospital and Institute of Psychiatry, University of Oslo, N-0407 Oslo, Norway. [6]Department of Clinical Neuroscience, HUBIN project, Karolinska Institutet and Hospital, SE-171 76 Stockholm, Sweden. [7]Department of Medical Genetics, University of Helsinki, 00014 Helsinki, Finland. [8]University of Iceland, Faculty of Medicine, 101 Reykjavik, Iceland.

Norwegian (0.21% allelic frequency), 4 out of 390 Finnish (0.51% allelic frequency) and 5 out of 590 Swedish (0.42% allelic frequency) samples. The variant was also observed in 1 out of 7,020 chromosomes by the National Heart, Lung, and Blood Institute (NHLBI) Exome Sequencing Project[21], and in 3 out of 31,714 chromosomes from a North American population using an exome SNP chip array (see Supplementary Information).

The effect of rs63750847-A is stronger when using elderly controls than when using general population controls (Table 1), which is a consequence of a greater frequency of the variant in the elderly. We estimate that the odds for carriers of rs63750847-A of reaching age 85 are 1.47-fold the odds of non-carriers.

As Alzheimer's disease is a common disease with a late onset, it is informative in association studies to use a control group that includes those who have reached old age without deficits in cognition. Therefore, we examined the frequency of rs63750847 in a control group of individuals who were cognitively intact at age 85, based on a score of 0 on the Cognitive Performance Scale (CPS), a seven-category hierarchical scale assessing cognitive function in the elderly (Supplementary Information). We found an enrichment (0.79%; OR = 7.52, $P = 6.92 \times 10^{-6}$; Table 1) of rs63750847-A in this group, consistent with a protective effect of rs63750847-A against Alzheimer's disease.

To study further the effect of the A673T substitution on cognitive decline in the elderly, we investigated cognitive function as measured with CPS in 41 carriers of A673T in the age range 80–100 as well as in 3,673 non-carriers. The Resident Assessment Instrument for Nursing Homes (RAI-NH), on which the CPS score is based, is applied on average three times per year in Icelandic Nursing Homes. Because the residency time in nursing homes in Iceland is on average 3–4 years, many determinations of CPS made at different times are available for most individuals (Supplementary Fig. 1). As expected, cognitive function declines slowly but steadily with age, both in carriers and non-carriers of A673T (Fig. 1). Analysing a total of 23,831 CPS scores for the 3,673 non-carriers of A673T without a diagnosis of Alzheimer's disease (average of 6.49 determinations per individual), and 262 CPS scores for the 41 carriers of A673T without a diagnosis of Alzheimer's disease (average of 6.39 determinations), we found on average a 1.03 unit difference between carriers and non-carriers across the 80–100 age range (Fig. 1; $P = 0.0021$), with the carriers having a score indicative of better conserved cognition. The fact that the cognitive function of non-carriers remained poorer than for carriers of A673T after removing known Alzheimer's disease cases suggests that the protective effect of A673T extends beyond the boundaries of the Alzheimer's disease phenotype.

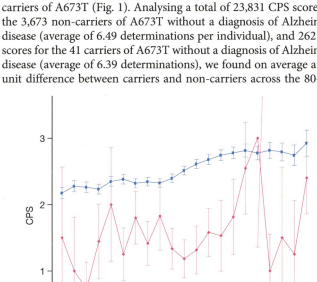


Figure 1 | Cognition measured by CPS as a function of age. Shown are CPS scores of carriers (red symbols) and non-carriers (blue symbols) of A673T as a function of age. Each symbol represents the average CPS score of individuals at the respective age (in years). Error bars represent ± 1 standard error. The jagged appearance of the graph for A673T carriers is due to the relatively small number of data points (262 in total, representing 41 individuals, as compared to 23,831 data points representing 3,673 A673T non-carriers). Individuals with a diagnosis of Alzheimer's disease were not included in the analysis.

The A673T substitution is located at position 2 in the amyloid-β peptide. Recently, an alanine to valine substitution at position 673 in APP (A673V) was reported as being recessive for Alzheimer's disease with very early onset in a single Italian pedigree[22,23]. Heterozygous carriers of A673V in this pedigree were unaffected. We identified three homozygous carriers of A673T in Icelandic samples, one of whom had died at age 88, with the other two currently living at age 67 and 83, respectively. None of these homozygous carriers had a history of dementia.

Our genetic data indicate that the A673T substitution in APP is protective against Alzheimer's disease. The proximity of A673T to the proteolytic site of BACE1 suggested to us that the variant might result in impaired BACE1 cleavage of APP in the A673T carriers.

To investigate the effect of A673T on proteolytic processing of APP, we followed the formation of extracellular APP fragments generated by APP processing at the β-site (sAPPβ) and α-site (sAPPα), respectively, as well as production of the amyloidogenic peptides $Aβ_{x-40}$ and $Aβ_{x-42}$, in 293T cells transfected with wild-type or mutant APP (Fig. 2). By western blot analysis of cell supernatants (Fig. 2a), we found that the A673T variant results in reduced production of sAPPβ with a slight apparent increase in production of sAPPα as compared to wild-type APP. We next confirmed these observations using a quantitative sandwich immunoassay approach (Fig. 2b). sAPPβ production from A673T was ~50% less than from wild-type APP, whereas sAPPα trended non-significantly towards an increase. We also found that the production of both amyloidogenic peptides $Aβ_{x-40}$ and $Aβ_{x-42}$ was ~40% less by the A673T variant than by wild-type APP (Fig. 2c, d). For comparison, we also analysed APP cleavage by the pathogenic A673V variant, which has previously been found to increase amyloidogenic processing of APP[22]. In contrast to A673T, the A673V substitution resulted in markedly increased APP processing at the β-site (Fig. 2a, b), decreased processing at the α-site (Fig. 2a, b), and greatly enhanced $Aβ_{x-40}$ and $Aβ_{x-42}$ production (Fig. 2c, d). For further reference, we also looked at $Aβ_{x-40}$ and $Aβ_{x-42}$ production by APP K670N/M671L, which has been reported to increase $Aβ_{x-40}$ and $Aβ_{x-42}$ production[24]. We confirmed that neither the A673T nor A673V substitution interfered with detection in the enzyme-linked immunosorbent assay (ELISA) (Supplementary Information). The change in the various APP cleavage products seen with A673T shows that this substitution reduces BACE1 cleavage of APP relative to wild-type APP, whereas A673V and K670N/M671L both markedly increase APP cleavage (Table 2). These results are consistent with the protective effect of A673T against Alzheimer's disease, as well as the dramatic phenotypic contrast between T and V substitution at the 673 site in APP. These data also illustrate clearly that position 673 of APP is capable of regulating proteolytic processing by BACE1.

To confirm these observations, we used an *in vitro* BACE1 cleavage assay to assess processing of a wild-type synthetic APP peptide substrate compared to a peptide bearing the A673T substitution. The A673T APP peptide was processed ~50% less efficiently than the wild-type substrate, supporting the conclusion that it codes a sub-optimal BACE1 cleavage site (see Supplementary Information). The substrate specificity of BACE1 has previously been investigated in synthetic model peptides, showing that amino acid substitutions at position 673 in APP can be tolerated[25]. Interestingly, although wild-type APP seems to be a relatively poor substrate for BACE1, most substitutions near the β-cleavage site result in an increased rate of cleavage of synthetic peptides[26]. However, consistent with our findings, a threonine substitution at position 673 of these APP peptide substrates leads to BACE1 cleavage rates that are 50-fold less than for a valine substitution at the same

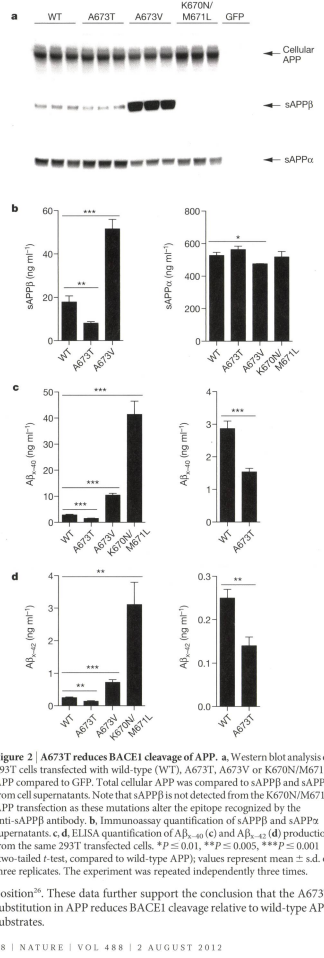


Figure 2 | A673T reduces BACE1 cleavage of APP. a, Western blot analysis of 293T cells transfected with wild-type (WT), A673T, A673V or K670N/M671L APP compared to GFP. Total cellular APP was compared to sAPPβ and sAPPα from cell supernatants. Note that sAPPβ is not detected from the K670N/M671L APP transfection as these mutations alter the epitope recognized by the anti-sAPPβ antibody. **b**, Immunoassay quantification of sAPPβ and sAPPα supernatants. **c**, **d**, ELISA quantification of $A\beta_{1-40}$ (**c**) and $A\beta_{1-42}$ (**d**) production from the same 293T transfected cells. $^{*}P \leq 0.01$, $^{**}P \leq 0.005$, $^{***}P \leq 0.001$ (two-tailed *t*-test, compared to wild-type APP); values represent mean ± s.d. of three replicates. The experiment was repeated independently three times.

position[26]. These data further support the conclusion that the A673T substitution in APP reduces BACE1 cleavage relative to wild-type APP substrates.

Table 2 | APP cleavage products from transfected 293T cells

	Wild type	A673T	A673V	K670N/M671L
sAPPβ	17.8 ± 2.9	8.1 ± 0.7	51.5 ± 4.4	N/A
sAPPα	527 ± 18	564 ± 21	476 ± 2	518 ± 33
$A\beta_{1-40}$	2.9 ± 0.2	1.5 ± 0.1	10.4 ± 0.7	41.4 ± 5.1
$A\beta_{1-42}$	0.25 ± 0.02	0.14 ± 0.02	0.72 ± 0.08	3.11 ± 0.69

All reported values are in ng ml^{-1}. APP cleavage products were quantified from supernatants from 293T cells transfected with wild-type, A673T, A673V or K670N/M671L APP. Values represent mean ± s.d. of three replicates from a single experiment. N/A, not applicable.

Our data show that position 673 of APP is critical for amyloidogenic processing of APP by BACE1. To our knowledge, A673T represents the first example of a sequence variant conferring strong protection against Alzheimer's disease. The strong protective effect of A673T also provides further proof of principle for the idea that reducing BACE1 cleavage of APP may protect against Alzheimer's disease. Furthermore, the fact that the A673T substitution also protects against cognitive decline in the elderly without Alzheimer's disease provides indirect support for the hypothesis that the pathogenesis of Alzheimer's disease and normal cognitive decline of the elderly may be shared, at least in part. We therefore propose that Alzheimer's disease may represent the extreme of the age-related decline in cognitive function.

METHODS SUMMARY

Patients with Alzheimer's disease were enrolled through the Memory Clinic at Landspitali University Hospital. Diagnosis of Alzheimer's disease was established according to NINCDS-ADRDA criteria or according to International Classification of Diseases, 10th revision (ICD-10) code F00 criteria. Cognitive function was assessed using the CPS, which is based on the Minimum Data Set for Nursing Homes. MDS 2.0, of the RAI by InterRAI[27].

Genotype data for A673T (rs63750847) were based on whole-genome sequence data generated from 1,795 Icelanders to a depth of at least ×10. Approximately 30 million markers (SNPs and indels) were imputed based on this set of individuals. Sequencing by synthesis was performed on Illumina GAIIx and HiSeq2000 instruments using previously described methods[18]. Long-range phasing of all chip-genotyped individuals was performed using previously described methods[15,28]. SNPs that were identified and genotyped through sequencing were imputed into all Icelanders who had been phased with long-range phasing using the same model used by IMPUTE[15]. Generation of *in silico* genotypes was performed by imputing genotypes into relatives of chip-genotyped individuals, using the fully phased imputed and chip-type genotypes of the available chip-typed individuals. Association testing was performed using logistic regression, matching controls to cases based on the informativeness of the imputed genotypes. Chip-typed samples were assayed with Illumina bead chips containing from 300,000 to 2,500,000 SNPs. SNPs that did not pass a rigorous quality control test were excluded. All samples with a call rate below 97% were also excluded.

Human APP695 cDNA was cloned into pRK vector and mutagenized using QuickChange site-directed mutagenesis kit (Stratagene), followed by transfection into 293T cells. APP cleavage products were assessed both by western blots and immunoassays. $A\beta_{1-40}$ and $A\beta_{1-42}$ peptides were measured from cell supernatants with sandwich ELISAs.

For further details, see Supplementary Information.

Received 14 March; accepted 6 June 2012.
Published online 11 July; corrected online 1 August 2012 (see full-text HTML for details).

1. Ferri, C. P. *et al.* Global prevalence of dementia: a Delphi consensus study. *Lancet* **366,** 2112–2117 (2005).
2. Plassman, B. L. *et al.* Prevalence of dementia in the United States: the aging, demographics, and memory study. *Neuroepidemiology* **29,** 125–132 (2007).
3. Qiu, C., Kivipelto, M. & von Strauss, E. Epidemiology of Alzheimer's disease: occurrence, determinants, and strategies toward intervention. *Dialogues Clin. Neurosci.* **11,** 111–128 (2009).
4. Reitz, C., Brayne, C. & Mayeux, R. Epidemiology of Alzheimer disease. *Nature reviews. Neurology* **7,** 137–152 (2011).
5. Glenner, G. G. & Wong, C. W. Alzheimer's disease: initial report of the purification and characterization of a novel cerebrovascular amyloid protein. *Biochem. Biophys. Res. Commun.* **120,** 885–890 (1984).
6. Masters, C. L. *et al.* Neuronal origin of a cerebral amyloid: neurofibrillary tangles of Alzheimer's disease contain the same protein as the amyloid of plaque cores and blood vessels. *EMBO J.* **4,** 2757–2763 (1985).
7. Zhang, Y. W., Thompson, R., Zhang, H. & Xu, H. APP processing in Alzheimer's disease. *Mol. Brain* **4,** 3 (2011).

8. Hussain, I. *et al.* Identification of a novel aspartic protease (Asp 2) as β-secretase. *Mol. Cell. Neurosci.* **14,** 419–427 (1999).
9. Sinha, S. *et al.* Purification and cloning of amyloid precursor protein β-secretase from human brain. *Nature* **402,** 537–540 (1999).
10. Vassar, R. *et al.* β-Secretase cleavage of Alzheimer's amyloid precursor protein by the transmembrane aspartic protease BACE. *Science* **286,** 735–741 (1999).
11. Yan, R. *et al.* Membrane-anchored aspartyl protease with Alzheimer's disease β-secretase activity. *Nature* **402,** 533–537 (1999).
12. Hardy, J. C. & Crook, R. APP mutations table. http://www.alzforum.org/res/com/mut/app/table1.asp (2010).
13. Cruts, M. Alzheimer's disease and frontotemporal dementia mutation database. http://www.molgen.ua.ac.be/ADMutations (2012).
14. St George-Hyslop, P. H. Molecular genetics of Alzheimer's disease. *Biol. Psychiatry* **47,** 183–199 (2000).
15. Cruchaga, C. *et al.* Rare variants in *APP*, *PSEN1* and *PSEN2* increase risk for AD in late-onset Alzheimer's disease families. *PLoS ONE* **7,** e31039 (2012).
16. Kong, A. *et al.* Detection of sharing by descent, long-range phasing and haplotype imputation. *Nature Genet.* **40,** 1068–1075 (2008).
17. Kong, A. *et al.* Fine-scale recombination rate differences between sexes, populations and individuals. *Nature* **467,** 1099–1103 (2010).
18. Marchini, J., Howie, B., Myers, S., McVean, G. & Donnelly, P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nature Genet.* **39,** 906–913 (2007).
19. Sulem, P. *et al.* Identification of low-frequency variants associated with gout and serum uric acid levels. *Nature Genet.* **43,** 1127–1130 (2011).
20. Peacock, M. L., Warren, J. T. Jr, Roses, A. D. & Fink, J. K. Novel polymorphism in the A4 region of the amyloid precursor protein gene in a patient without Alzheimer's disease. *Neurology* **43,** 1254–1256 (1993).
21. Exome Variant Server. NHLBI Exome Sequencing Project (ESP). http://evs.gs.washington.edu/EVS/ (2012).
22. Di Fede, G. *et al.* A recessive mutation in the APP gene with dominant-negative effect on amyloidogenesis. *Science* **323,** 1473–1477 (2009).
23. Giaccone, G. *et al.* Neuropathology of the recessive A673V *APP* mutation: Alzheimer disease with distinctive features. *Acta Neuropathol.* **120,** 803–812 (2010).
24. Citron, M. *et al.* Mutation of the β-amyloid precursor protein in familial Alzheimer's disease increases β-protein production. *Nature* **360,** 672–674 (1992).
25. Gruninger-Leitch, F., Schlatter, D., Kung, E., Nelbock, P. & Dobeli, H. Substrate and inhibitor profile of BACE (β-secretase) and comparison with other mammalian aspartic proteases. *J. Biol. Chem.* **277,** 4687–4693 (2002).
26. Tomasselli, A. G. *et al.* Employing a superior BACE1 cleavage sequence to probe cellular APP processing. *J. Neurochem.* **84,** 1006–1017 (2003).
27. Morris, J. N. *et al.* MDS Cognitive Performance Scale. *J. Gerontol.* **49,** M174–M182 (1994).
28. Holm, H. *et al.* A rare variant in MYH6 is associated with high risk of sick sinus syndrome. *Nature Genet.* **43,** 316–320 (2011).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We would like to thank the NHLBI GO Exome Sequencing Project and its ongoing studies, which produced and provided exome variant calls for comparison: the Lung GO Sequencing Project (HL-102923), the WHI Sequencing Project (HL-102924), the Broad GO Sequencing Project (HL-102925), the Seattle GO Sequencing Project (HL-102926) and the Heart GO Sequencing Project (HL-103010).

Author Contributions The study was designed and results were interpreted by T.J., J.K.A., H.S., R.J.W. and K.S. Sequence data analysis was carried out by T.J., S.S., P.S., A.K., T.B., R.R.G., T.W.B. and D.G. Subject recruitment, phenotype analysis and biological material collection was organized and carried out by J.S., P.V.J., S.B., G.B., O.A.A., E.G.J. and A.P. Sequencing and genotyping was supervised by J.H., O.T.M. and U.T. Cell line experiments and BACE1 cleavage assays were carried out and analysed by J.K.A., J.M., K.H., Y. Lu, Y. Liu, A.G. and R.J.W. The paper was drafted by T.J., J.K.A., R.J.W. and K.S. All authors contributed to the final version of the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the full-text HTML version of the paper at www.nature.com/nature. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to K.S. (kstefans@decode.is).

LETTER

doi:10.1038/nature11284

Dissecting the genomic complexity underlying medulloblastoma

A list of authors and their affiliations appears at the end of the paper

Medulloblastoma is an aggressively growing tumour, arising in the cerebellum or medulla/brain stem. It is the most common malignant brain tumour in children, and shows tremendous biological and clinical heterogeneity[1]. Despite recent treatment advances, approximately 40% of children experience tumour recurrence, and 30% will die from their disease. Those who survive often have a significantly reduced quality of life. Four tumour subgroups with distinct clinical, biological and genetic profiles are currently identified[2,3]. WNT tumours, showing activated wingless pathway signalling, carry a favourable prognosis under current treatment regimens[4]. SHH tumours show hedgehog pathway activation, and have an intermediate prognosis[2]. Group 3 and 4 tumours are molecularly less well characterized, and also present the greatest clinical challenges[2,3,5]. The full repertoire of genetic events driving this distinction, however, remains unclear. Here we describe an integrative deep-sequencing analysis of 125 tumour–normal pairs, conducted as part of the International Cancer Genome Consortium (ICGC) PedBrain Tumor Project. Tetraploidy was identified as a frequent early event in Group 3 and 4 tumours, and a positive correlation between patient age and mutation rate was observed. Several recurrent mutations were identified, both in known medulloblastoma-related genes (*CTNNB1*, *PTCH1*, *MLL2*, *SMARCA4*) and in genes not previously linked to this tumour (*DDX3X*, *CTDNEP1*, *KDM6A*, *TBR1*), often in subgroup-specific patterns. RNA sequencing confirmed these alterations, and revealed the expression of what are, to our knowledge, the first medulloblastoma fusion genes identified. Chromatin modifiers were frequently altered across all subgroups. These findings enhance our understanding of the genomic complexity and heterogeneity underlying medulloblastoma, and provide several potential targets for new therapeutics, especially for Group 3 and 4 patients.

As a first phase of the International Cancer Genome Consortium (ICGC) PedBrain Tumor Project (http://www.pedbraintumor.org), we have collected matched tumour and germline samples from 125 medulloblastoma patients aged from 0 to 17 years (Supplementary Table 1). Whole-genome sequencing (WGS, $n = 39$) and whole-exome sequencing (WES, $n = 21$) were applied to a 'discovery' set, with a custom-capture approach used to sequence 2,734 genes in an additional 'replication' set ($n = 65$). All tumour samples were obtained at primary diagnosis, before adjuvant therapy, and the distribution of molecular subgroups was similar across cohorts (Supplementary Fig. 1).

Investigation of genome-wide somatic mutation allele frequencies identified several cases with a clear peak at approximately 25%, rather than the expected approximately 50% allele frequency for early, heterozygous events (Fig. 1a). Analysis of coverage depth and allele frequencies in regions of copy-number change ruled out stromal contamination, but rather indicated a tetraploid baseline in the tumour genome (Fig. 1b). Predicted ploidy status was confirmed by fluorescence *in situ* hybridization (FISH) using multiple centromeric probes in 17 out of 18 cases analysed (Fig. 1a). The extremely low fraction of mutations at approximately 50% allele frequency indicates that genome duplication occurred very early during tumorigenesis. Some cases probably went through even higher polyploidy states before reaching an approximately 4n baseline (for example ICGC_MB45, displaying 4n chromosomes with 4:0 or 3:1 allele ratios; Supplementary Fig. 2). Across the discovery set, tetraploidy was most commonly observed in Group 3 (7 out of 13, 54%) and Group 4 tumours (8 out of 20, 40%), followed by SHH (4 out of 14, 29%) and WNT tumours (1 out of 7, 14%). Interestingly, the four tetraploid SHH tumours all harboured *TP53* mutations and also displayed chromothripsis[6]. Tetraploid Group 3 and 4 tumours showed significantly more large-scale copy number alterations compared with diploid cases (median 10 changes per tumour in tetraploid versus 4 per tumour in diploid cases, $P = 0.008$, two-tailed Mann–Whitney U-test; Supplementary Fig. 3). Thus, tetraploidy followed by genomic instability may be an early driving event in a large proportion of Group 3 and 4 medulloblastomas, which pose a significant clinical challenge due to their dismal prognosis and lack of targeted treatment options. Novel classes of drugs such as mitotic checkpoint kinase or kinesin inhibitors, which target the maintenance of tetraploidy through successive cell divisions, may therefore represent a rational therapeutic strategy in these cases[7,8]. The value of tetraploidy as a prognostic marker also requires further investigation.

The average somatic mutation rate in the WGS cohort was 0.52 per megabase (Mb), with an average of 10.3 non-synonymous coding single-nucleotide variants (SNVs) in the discovery cohort (Supplementary Table 2). This is slightly higher than previously reported for medulloblastoma[9], possibly due to improved coverage and technical sensitivity, but considerably lower than in deep-sequenced adult tumours, for example[10,11]. There were significantly fewer transitions in the somatic alterations compared with germline variation ($P = 4.6 \times 10^{-7}$, Wilcoxon rank-sum test; Supplementary Fig. 4). All coding somatic SNVs identified in the combined cohort are listed in Supplementary Table 3.

We identified a positive correlation between genome-wide mutation rate and patient age, as previously reported for coding mutations[9] ($r^2 = 0.35$, $P = 7.8 \times 10^{-5}$ Pearson's product–moment correlation; Fig. 1c). Intriguingly, this association was more pronounced in diploid tumours ($r^2 = 0.52$, $P = 3 \times 10^{-5}$), and virtually absent in tetraploid cases ($r^2 = 0.04$, $P = 0.5$) (Supplementary Fig. 5a, b). A similar trend was observed for non-synonymous mutations across the discovery cohort (Supplementary Fig. 5c). Coverage level did not correlate with mutation rate (Supplementary Fig. 5d). One explanation may be that all medulloblastomas originate during embryogenesis, with some tumours needing to accumulate more genetic 'hits' before becoming symptomatic. Alternatively, tumours arising in older patients may derive from more differentiated cells that require a greater number of alterations to undergo malignant transformation. Investigation of additional tumours from older patients may help to clarify this.

Five SHH tumours harbouring *TP53* mutations, including three previously described Li–Fraumeni syndrome (LFS)-associated tumours with germline mutations[6], one newly identified LFS case (ICGC_MB23), and one somatically mutated tumour (ICGC_MB34), had significantly more mutations than the remaining cases, both genome wide (mean 1.1 per Mb versus 0.43 per Mb, $P = 4.5 \times 10^{-6}$; two-tailed t-test) and for

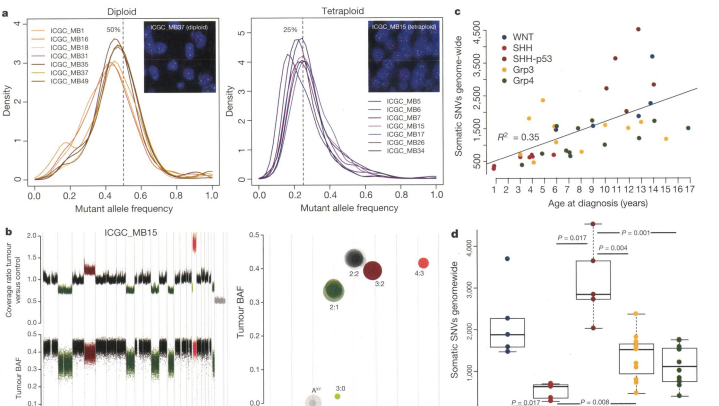


Figure 1 | Tetraploidy is a frequent early event in medulloblastoma tumorigenesis, and mutation rates vary with age and subgroup. **a**, Distributions of genome-wide somatic mutation allele frequencies (the proportion of sequence reads supporting a mutation) for diploid tumours (with a peak at ~50% for heterozygous events, $n = 7$) and tetraploid cases (with a peak at ~25%, $n = 7$). Insets show centromeric FISH for chromosomes 1 (red) and 11 (green), confirming the predicted ploidy status. **b**, Top left, rescaled tumour:germline coverage ratio, indicating copy-number gains (red) or losses (green). Bottom left, B-allele frequency (BAF) in the tumour at SNP positions which are heterozygous in the germ line. Right, genome alteration print (GAP) of segmented copy number and allele frequency profiles. Chromosomes with predicted 3:0/2:1/3:2 allele ratios show a BAF of approximately 0/0.33/0.4 and coverage ratios of approximately 0.75/0.75/1.25. Owing to random sampling, the 2:2 allele ratio is slightly below 0.5. **c**, Genome-wide somatic mutation rates are positively correlated with patient age ($n = 39$). Grp, Group. **d**, Distribution of somatic mutation rates by tumour subgroup ($n = 39$). *P* values are according to a Wilcoxon rank-sum test with Bonferroni correction. SHH-p53, SHH-subgroup tumours harbouring a somatic or germline *TP53* mutation.

non-synonymous changes (mean 23 versus 8.8, $P = 2.6 \times 10^{-6}$). Interestingly, the WNT subgroup, which typically shows a good prognosis and few copy-number changes, had the next highest mutation rate (Fig. 1d).

Forty-one somatic, coding, small insertions/deletions (Indels) were identified across the cohort, with an average of 0.4 coding Indels per case in the discovery set (range 0–2; Supplementary Table 4). Some genes, however, were more commonly affected by Indels than SNVs. For example, frameshift Indels in *PTCH1* were detected in 6 out of 125 cases, whereas only 2 SNVs were observed. Recurrent Indels were also seen in the chromatin modifiers *MLL2*, *KDM6A* (3 cases each) and *BCOR* (2 cases).

In contrast to another paediatric brain tumour, glioblastoma, in which we recently identified frequently recurrent hotspot mutations[12], the majority of mutated genes in this study were unique to a single case (587 out of 760 non-synonymous SNVs in the 125 cases, 77%), demonstrating the pronounced genetic heterogeneity of medulloblastoma. Twenty-five of these singleton mutations, and 53 SNVs in total, were at positions listed in the COSMIC database of somatic alterations in tumours (available at http://www.sanger.ac.uk/genetics/CGP/cosmic/), suggesting a rare but important contribution of many known cancer genes in medulloblastoma (Supplementary Table 5). Only 8 genes were somatically altered in more than 3% of the whole series: *CTNNB1* (15 cases, 12%); *DDX3X* (10 cases, 8%); *PTCH1* (8 cases, 6%), *SMARCA4* (6 cases, 5%), *MLL2* (6 cases, 5%), *TP53* (somatically mutated in 5 cases, 4%), *KDM6A* (5 cases, 4%) and *CTDNEP1* (4 cases, 3%) (Fig. 2). These were also the only genes found to be significantly altered upon analysis of the combined cohort with MutSig, an algorithm testing whether the observed mutations in a gene are not simply a consequence of random background mutation processes. It takes into account gene length and composition, silent to non-silent mutation ratios, and other factors (see https://confluence.broadinstitute.org/display/CGATools/MutSig; Supplementary Table 6). Large-scale copy-number changes known to be associated with medulloblastoma, such as formation of an isodicentric 17q and losses of 10q/9q/X[13–15], were more frequently recurrent than SNVs (Supplementary Fig. 6a–e).

Many alterations were enriched in specific medulloblastoma subgroups. For example, all of the WNT tumours (15 out of 15) harboured a mutation in *CTNNB1*, and 13 out of 15 displayed loss of one copy of chromosome 6 (or acquired uniparental disomy in one case), alterations which have previously been associated with this subgroup[4,13,15]. Mutations in *DDX3X* were also clearly enriched in WNT tumours (adjusted $P = 7.06 \times 10^{-6}$, two-tailed Fisher's exact test with a Bonferroni correction), and these mutations were clustered within the helicase domain (Supplementary Fig. 7a). Three were localized at the RNA-binding surface of the protein and three were predicted to disrupt the closed (RNA-binding) conformation (Supplementary Fig. 7b). The remainder were predicted to disrupt indirectly either the positive charge on the RNA-binding surface ($n = 2$) or the folding of the closed form ($n = 2$). No truncating mutations were found,

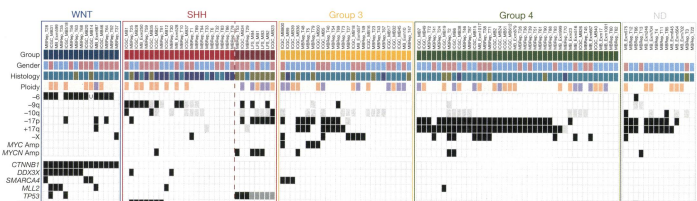


Figure 2 | Subgroup specificity of common genetic alterations. Summary of clinical data and recurrent alterations in the combined cohort ($n = 125$). Genes which were found to be significantly mutated by MutSig analysis were included. UPD, uniparental disomy; ND, no material available for conclusive molecular subgroup assignment.

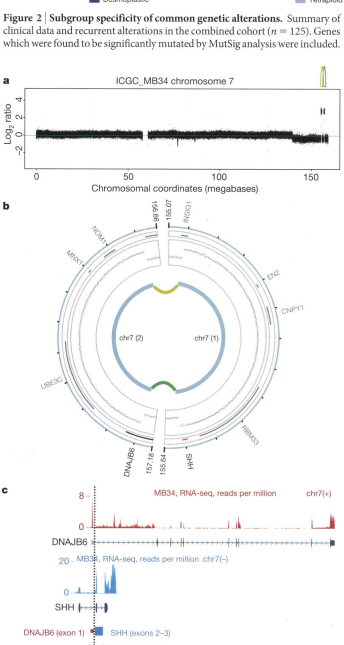


Figure 3 | Identification of novel fusion genes in medulloblastoma. a, Read-depth plot with $\log_2$ tumour:germline coverage ratio showing alterations on chromosome 7 in ICGC_MB34. Lines indicate connected segments. **b**, Schematic of the rearrangement. **c**, Details of the *SHH* fusion gene structure and support for its expression, derived from RNA sequencing data. aa, amino acids.

indicating an alteration rather than simply a loss of function. *DDX3X* has recently been proposed to have an oncogenic role[10,11], although its exact function in tumorigenesis remains to be determined.

As anticipated from previous studies[13,16], SHH tumours frequently showed loss of the whole of chromosome arm 9q, as well as alterations in key hedgehog-pathway signalling molecules (for example, *PTCH1*, altered in 8 cases; *MYCN*, amplified in 5 cases; and *SMO*, mutated in ICGC_MB12).

The most frequently mutated gene in Group 3 tumours was *SMARCA4* (3 out of 26 cases). As with *DDX3X*, these mutations were clustered in the helicase domain (Supplementary Fig. 7a). As noted above, tetraploidy was also a common event in this subgroup and in Group 4 tumours. Recurrent truncating mutations in *KDM6A* (on chromosome X, which frequently shows copy-number loss in female Group 3 and 4 medulloblastoma patients; also known as *UTX*), encoding a histone 3 lysine 27 (H3K27) demethylase, were also seen in Group 4 (4 out of 40, 10%), indicating a tumour-suppressive role in this subgroup, as previously described for other cancers[17]. *CTDNEP1* (a homologue of the *Xenopus* gene *dullard*), was also affected by truncating alterations in four tumours. In three of these cases, the mutation was accompanied by loss of the wild-type allele through isodicentric 17q formation. This gene, encoding a nuclear envelope phosphatase, was shown in *Xenopus* to have roles in BMP signalling and neural development[18]. In mammalian cells it is involved in the lipin activation pathway, regulating nuclear membrane biogenesis and production of diacylglycerol[19,20]. Given the high frequency of isodicentric 17q in medulloblastoma, genetic targets on this chromosome have long been sought after. *CTDNEP1* may be a good candidate for one of the medulloblastoma tumour suppressors on 17p.

Aside from these subgroup-enriched events, a commonly recurring theme across all medulloblastomas is alterations in genes involved in chromatin modification. Some point mutations and DNA copy number alterations in this pathway have previously been implicated in medulloblastoma[9,21]. Overall, 45 out of 125 cases (36%) harboured a mutation in a gene categorized under the Gene Ontology term 'Chromatin Modification' (GO:0015168, Supplementary Fig. 6f, g).

We recently described an enrichment of catastrophic DNA rearrangements ('chromothripsis') in *TP53*-mutated SHH medulloblastomas[6]. Three new *TP53*-mutant SHH tumours were identified in this

study: ICGC_MB23 (germline mutation), MBRep_T29 and MBRep_T53 (somatic mutations). Two of these, ICGC_MB23 and MBRep_T53, showed complex genomic rearrangements indicative of the chromothripsis model (Supplementary Fig. 8)[22].

Deep sequencing also allowed fine mapping of two amplicons on chromosome 7 in ICGC_MB34 (a SHH tumour with a somatic *TP53* mutation, relating to MB2034 in ref. 6). One amplicon included the entire *SHH* gene, whereas the second disrupted *DNAJB6*, such that its first exon was juxtaposed to *SHH* (Fig. 3a, b). RNA sequencing further revealed a novel fusion transcript, not expected from the DNA data, containing the first exon of *DNAJB6* and exons 2 and 3 of *SHH*. The first exon of *SHH* was skipped, resulting in a predicted amino-terminally truncated SHH protein (Fig. 3c). Expression of SHH was extremely high in this case, although virtually absent in 301 other medulloblastomas (Supplementary Fig. 9a). Predicted DNA and RNA junctions were validated by PCR (Supplementary Fig. 9b).

Several additional in-frame gene fusions were identified by large insert mate-pair sequencing, which gives better resolution for structural variant detection. ICGC_MB18, for example, carried an intrachromosomal translocation resulting in a fusion between *LCLAT1* and *ERBB4*, the latter of which has previously been associated with medulloblastoma oncogenesis[23] (Supplementary Fig. 9c–f). In ICGC_MB6, a complex rearrangement of fragments from chromosomes 1 and 17 produced a fusion between *MLLT6* and *MRPL45*, a mitochondrial ribosomal protein, resulting in strong overexpression of the latter (Supplementary Fig. 10a–c). These findings indicate that gene fusions involving well-established medulloblastoma oncogenes may have a more important role in medulloblastoma than previously recognized, and warrant further investigation.

High-coverage, strand-specific RNA sequencing of 28 cases allowed us to determine the proportion of DNA SNVs that were observable in the transcriptome (Supplementary Tables 3 and 4). Overall, 129 out of 268 (48%) non-synonymous mutations in the DNA were also detectable at the RNA level. A further 38% (101 out of 268) resided in genes expressed at extremely low abundance (reads per kilobase of exon model per million mapped reads (RPKM) < 1). Thus, the fraction of expressed mutations is even smaller than the already low number of DNA alterations, supporting the hypothesis that very few driving hits are needed to generate this paediatric tumour. It may also be the case that some mutations required for tumour initiation are not essential for later tumour cell maintenance.

RNA sequencing further revealed monoallelic expression of a heterozygous mutation in *TBR1*, producing a p.G275C change, which was also seen in a previous study[9] (Supplementary Fig. 11a). *TBR1* encodes a T-box transcription factor involved in brain development[24]. This gene, and a second family member, *EOMES* (or *TBR2*), clearly showed subgroup-specific differential expression (Fig. 4a). Sequencing of *TBR1* exon 2 in a further 85 medulloblastomas revealed one additional case with an identical mutation. All three mutated tumours were in Group 4. Gene expression was also strongly correlated with DNA methylation for both *TBR1* and *EOMES* (Fig. 4b, c and Supplementary Fig. 11b, c), and expression of *TBR1* and *EOMES* is inversely correlated

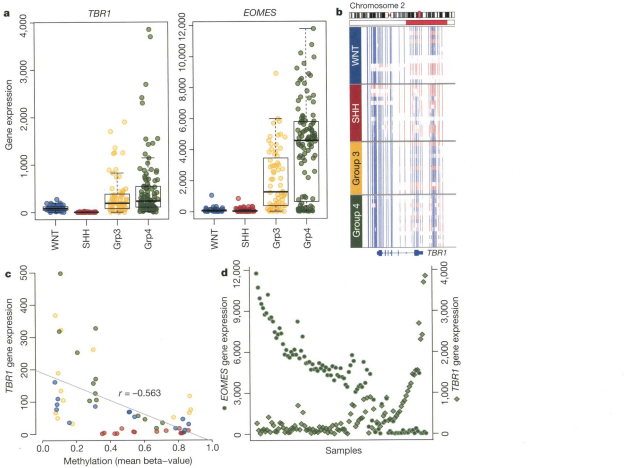


Figure 4 | Integration of mutation, expression and methylation data shows differential regulation of *TBR1* and *EOMES* in medulloblastoma. **a**, Microarray data showing clear differences in *TBR1* and *EOMES* expression between medulloblastoma subgroups ($n = 301$). **b**, DNA methylation of *TBR1* ($n = 54$), ranging from low (blue) to high (red). Horizontal red bar indicates the region used for correlation analysis in **c**. **c**, Expression of *TBR1* is tightly correlated with gene methylation ($n = 54$; Pearson's correlation values, r). SHH tumours show high methylation and virtually no expression, whereas WNT, Group 3 and Group 4 tumours display a more varied pattern. **d**, Expression levels of *TBR1* (diamonds) and *EOMES* (circles) are inversely related in Group 4 tumours ($n = 104$).

in Group 4 tumours (Fig. 4d), giving subsets that are either *TBR1*-methylated and EOMES^{hi} or *EOMES*-methylated and TBR1^{hi} (Supplementary Fig. 11d, e). These two genes are markers for different stages of neuronal lineage commitment, suggesting possible differences in cell-of-origin or differentiation within Group 4 subpopulations[25].

This large, integrative genomics study has provided a detailed insight into new mechanisms contributing to medulloblastoma tumorigenesis and disclose novel targets for therapeutic approaches, especially for Group 3 and 4 patients. The molecular subgroup-related enrichment of many alterations highlights the importance of considering this distinguishing factor in research, trial design and clinical practice.

METHODS SUMMARY

All patient material was collected after receiving informed consent according to ICGC guidelines and as approved by the institutional review board of contributing centres. Tumour subgrouping was based on gene expression profiling or immunohistochemical analysis as described in ref. 5.

Next generation sequencing was performed using Illumina technologies. Mean DNA sequence coverage was 35-fold for whole-genome cases (range 26–56×), whereas mean on-target coverage in the whole-exome and replication cohorts was 68-fold (74% of targets above 20× for whole exome, 66% for the replication cohort). Exome capture was carried out with Agilent SureSelect (Human All Exon 50 Mb and XT Custom Library) in-solution reagents. Sequence data were aligned to the hg19 human reference genome assembly; duplicate and non-uniquely mapping reads were excluded. Tumour ploidy was predicted from sequencing data by a novel approach integrating copy number aberrations with allele frequencies. A subset of sequence variants were validated using PCR and Sanger sequencing. Verification rates were 95% (128 out of 135) for SNVs and 100% (14 out of 14) for Indels (Supplementary Tables 3 and 4). A complete description of the materials and methods is provided in the Supplementary Information.

Received 3 February; accepted 6 June 2012.
Published online 25 July 2012.

1. Louis, D. N. *et al.* The 2007 WHO classification of tumours of the central nervous system. *Acta Neuropathol.* **114,** 97–109 (2007).
2. Kool, M. *et al.* Molecular subgroups of medulloblastoma: an international meta-analysis of transcriptome, genetic aberrations, and clinical data of WNT, SHH, Group 3, and Group 4 medulloblastomas. *Acta Neuropathol.* **123,** 473–484 (2012).
3. Taylor, M. D. *et al.* Molecular subgroups of medulloblastoma: the current consensus. *Acta Neuropathol.* **123,** 465–472 (2012).
4. Clifford, S. C. *et al.* Wnt/Wingless pathway activation and chromosome 6 loss characterise a distinct molecular sub-group of medulloblastomas associated with a favourable prognosis. *Cell Cycle* **5,** 2666–2670 (2006).
5. Northcott, P. A. *et al.* Medulloblastoma comprises four distinct molecular variants. *J. Clin. Oncol.* **29,** 1408–1414 (2011).
6. Rausch, T. *et al.* Genome sequencing of pediatric medulloblastoma links catastrophic DNA rearrangements with TP53 mutations. *Cell* **148,** 59–71 (2012).
7. Rello-Varona, S. *et al.* Preferential killing of tetraploid tumor cells by targeting the mitotic kinesin Eg5. *Cell Cycle* **8,** 1030–1035 (2009).
8. Vitale, I. *et al.* Inhibition of Chk1 kills tetraploid tumor cells through a p53-dependent pathway. *PLoS ONE* **2,** e1337 (2007).
9. Parsons, D. W. *et al.* The genetic landscape of the childhood cancer medulloblastoma. *Science* **331,** 435–439 (2011).
10. Stransky, N. *et al.* The mutational landscape of head and neck squamous cell carcinoma. *Science* **333,** 1157–1160 (2011).
11. Wang, L. *et al. SF3B1* and other novel cancer genes in chronic lymphocytic leukemia. *N. Engl. J. Med.* **365,** 2497–2506 (2011).
12. Schwartzentruber, J. *et al.* Driver mutations in histone H3.3 and chromatin remodelling genes in paediatric glioblastoma. *Nature* **482,** 226–231 (2012).
13. Kool, M. *et al.* Integrated genomics identifies five medulloblastoma subtypes with distinct genetic profiles, pathway signatures and clinicopathological features. *PLoS ONE* **3,** e3088 (2008).
14. Pfister, S. *et al.* Outcome prediction in pediatric medulloblastoma based on DNA copy-number aberrations of chromosomes 6q and 17q and the *MYC* and *MYCN* loci. *J. Clin. Oncol.* **27,** 1627–1636 (2009).
15. Thompson, M. C. *et al.* Genomics identifies medulloblastoma subgroups that are enriched for specific genetic alterations. *J. Clin. Oncol.* **24,** 1924–1931 (2006).
16. Pietsch, T. *et al.* Medulloblastomas of the desmoplastic variant carry mutations of the human homologue of *Drosophila* patched. *Cancer Res.* **57,** 2085–2088 (1997).
17. van Haaften, G. *et al.* Somatic mutations of the histone H3K27 demethylase gene *UTX* in human cancer. *Nature Genet.* **41,** 521–523 (2009).
18. Satow, R., Kurisaki, A., Chan, T. C., Hamazaki, T. S. & Asashima, M. Dullard promotes degradation and dephosphorylation of BMP receptors and is required for neural induction. *Dev. Cell* **11,** 763–774 (2006).
19. Han, S. *et al.* Nuclear envelope phosphatase 1-regulatory subunit 1 (formerly TMEM188) is the metazoan Spo7p ortholog and functions in the lipin activation pathway. *J. Biol. Chem.* **287,** 3123–3137 (2012).
20. Kim, Y. *et al.* A conserved phosphatase cascade that regulates nuclear membrane biogenesis. *Proc. Natl Acad. Sci. USA* **104,** 6596–6601 (2007).
21. Northcott, P. A. *et al.* Multiple recurrent genetic events converge on control of histone lysine methylation in medulloblastoma. *Nature Genet.* **41,** 465–472 (2009).
22. Stephens, P. J. *et al.* Massive genomic rearrangement acquired in a single catastrophic event during cancer development. *Cell* **144,** 27–40 (2011).
23. Gilbertson, R. J., Perry, R. H., Kelly, P. J., Pearson, A. D. J. & Lunec, J. Prognostic significance of HER2 and HER4 coexpression in childhood medulloblastoma. *Cancer Res.* **57,** 3272–3280 (1997).
24. Hevner, R. F. *et al.* Tbr1 regulates differentiation of the preplate and layer 6. *Neuron* **29,** 353–366 (2001).
25. Englund, C. *et al.* Pax6, Tbr2, and Tbr1 are expressed sequentially by radial glia, intermediate progenitor cells, and postmitotic neurons in developing neocortex. *J. Neurosci.* **25,** 247–251 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank GATC Biotech AG for sequencing services. For technical support and expertise we thank: B. Haase, D. Pavlinic, B. Baying, M. Wahlers, R. Lück, I. Kutschera, K. Schlangen, M. Metsger, K. Schulz, A. Nürnberger, A. Kovacsovics, M. Linser, J. C. Lindsey, S. Bailey, D. M. Pearson, the EMBL Genomics Core Facility, the EMBL High-performance Computing Core Facility and the DKFZ Genomics and Proteomics Core Facility. This work was principally supported by the PedBrain Tumor Project contributing to the International Cancer Genome Consortium (ICGC PedBrain Tumor Project, http://www.pedbraintumor.org/), funded by German Cancer Aid (109252) and the German Federal Ministry of Education and Research (BMBF, NGFN^{plus} #01GS0883). Additional support came from the German Cancer Research Center–Heidelberg Center for Personalized Oncology (DKFZ-HIPO), the Max Planck Society, the Pediatric Brain Tumor Foundation, the Italian Neuroblastoma Foundation and the Samantha Dickson Brain Tumour Trust. This study included samples provided by the UK Children's Cancer and Leukaemia Group (CCLG) as part of CCLG-approved biological study BS-2007-04.

Author Contributions D.T.W.J., M.Su., A.M.S., H.-J.W., S.B., S.P., H.C., E.P., L.S., A.W., S.H., T.T., B.R., C.C.B., M.Sch., C.v.K., V.B., R.V., S.Wo., S.Wi. and J.F. performed and/or coordinated experimental work. N. Jäger, D.T.W.J., M.K., T.Z., B.H., M.Su, T.J.P., V.Ho., T.R., H.-J.W., J.W., M.A., V.Am, M.Z., Q.W., B.L., V.Ast, C.L., J.E., R.K., P.v.S., J.K., D.Sh., M.J.B., R.B.R. and P.A.N. performed data analysis. Y.-J.C., M.Ry., M.Ne., S.C., G.P.T., U.S., V.Ha., N.G., Y.-J.K., C.M., W.R., A.U., C.H.-M., T.M., A.E.K., A.v.D., O.W., E.M., J.R., M.E., M.U.S., M.C.F., M.H., N. Jabado, S.R., A.O.v.B., D.W., S.C.C., M.G.M., V.P.C., W.S., G.R., M.D.T. and A.K. collected data and provided patient materials. D.T.W.J., N. Jäger, D.St., M.K., V.Ho., H.W., R.E., S.M.P. and P.L. prepared the initial manuscript and figures. U.D.W., H.L., B.B., G.R., M.M., S.L.P., M.-L.Y., J.O.K., R.E., A.K., S.M.P. and P.L. provided project leadership. All authors contributed to the final manuscript.

Author Information Short-read sequencing data have been deposited at the European Genome-phenome Archive (EGA, http://www.ebi.ac.uk/ega/) hosted by the EBI, under accession number EGAS00001000215. Reprints and permissions information is available at www.nature.com/reprints. This paper is distributed under the terms of the Creative Commons Attribution-Non-Commercial-Share Alike licence, and is freely available to all readers at www.nature.com/nature. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to R.E. (r.eils@dkfz-heidelberg.de), S.M.P (s.pfister@dkfz-heidelberg.de) or P.L. (m.macleod@dkfz-heidelberg.de).

David T. W. Jones[1]*, Natalie Jäger[2]*, Marcel Kool[1], Thomas Zichner[3], Barbara Hutter[2], Marc Sultan[3], Yoon-Jae Cho[5], Trevor J. Pugh[6], Volker Hovestadt[1], Adrian M. Stütz[3], Tobias Rausch[3], Hans-Jörg Warnatz[4], Marina Ryzhova[8], Sebastian Bender[1], Dominik Sturm[1], Sabrina Pleier[1], Huriye Cin[1], Elke Pfaff[1], Laura Sieber[1], Andrea Wittmann[1], Marc Remke[1], Hendrik Witt[1,9], Sonja Hutter[1], Theophilos Tzaridis[1], Joachim Weischenfeldt[3], Benjamin Raeder[2], Meryem Avci[2], Vyacheslav Amstislavskiy[4], Marc Zapatka[7], Ursula D. Weber[7], Qi Wang[2], Bärbel Lasitschka[10], Cynthia C. Bartholomae[11], Manfred Schmidt[11], Christof von Kalle[11], Volker Ast[12], Christa Lawerenz[12], Jürgen Eils[12], Rolf Kabbe[2], Vladimir Benes[13], Peter van Sluis[14], Jan Koster[14], Richard Volckmann[14], David Shih[15], Matthew J. Betts[15], Robert B. Russell[16], Simona Coco[17], Gian Paolo Tonini[17], Ulrich Schüller[18], Volkmar Hans[19], Norbert Graf[20], Yoo-Jin Kim[21], Camelia Monoranu[22], Wolfgang Roggendorf[22], Andreas Unterberg[23], Christel Herold-Mende[23], Till Milde[5,24], Andreas E. Kulozik[5], Andreas von Deimling[25,26], Olaf Witt[5,24], Eberhard Maass[27], Jochen Rössler[28], Martin Ebinger[28], Martin U. Schuhmann[29], Michael C. Frühwald[21], Martin Hasselblatt[30], Nada Jabado[31], Stefan Rutkowski[32], André O. von Bueren[34], Dan Williamson[33], Steven C. Clifford[33], Mark G. McCabe[35,37], V. Peter Collins[37], Stephan Wolf[10], Stefan Wiemann[10,38], Hans Lehrach[4], Benedikt Brors[3], Wolfram Scheurlen[29], Jörg Felsberg[20], Guido Reifenberger[20] Paul A. Northcott[15], Michael D. Taylor[41], Matthew Meyerson[42,43], Scott L. Pomeroy[5,43], Marie-Laure Yaspo[4], Jan O. Korbel[3], Andrey Korshunov[25,26], Roland Eils[2,44,45]*, Stefan M. Pfister[1,9]* & Peter Lichter[1]*

[1]Division of Pediatric Neurooncology, German Cancer Research Center (DKFZ), Im Neuenheimer Feld 280, Heidelberg 69120, Germany. [2]Division of Theoretical Bioinformatics, German Cancer Research Center (DKFZ), Im Neuenheimer Feld 280,

Heidelberg 69120, Germany. [3]European Molecular Biology Laboratory (EMBL), Meyerhofstrasse 1, Heidelberg 69117, Germany. [4]Max Planck Institute for Molecular Genetics, Ihnestrasse 63-73, Berlin 14195, Germany. [5]Division of Child Neurology, Stanford University, 750 Welch Road, Palo Alto, California 94304, USA. [6]Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA. [7]Division of Molecular Genetics, German Cancer Research Center (DKFZ), Im Neuenheimer Feld 280, Heidelberg 69120, Germany. [8]Department of Neuropathology, NN Burdenko Neurosurgical Institute, 4th Tverskaya-Yamskaya 16, Moscow 125047, Russia. [9]Department of Pediatric Oncology, Hematology & Immunology, Heidelberg University Hospital, Im Neuenheimer Feld 430, Heidelberg 69120, Germany. [10]Genomics and Proteomics Core Facility, German Cancer Research Center (DKFZ), Im Neuenheimer Feld 280, Heidelberg 69120, Germany. [11]Division of Translational Oncology, German Cancer Research Center (DKFZ) and National Center for Tumor Diseases (NCT), Im Neuenheimer Feld 460, Heidelberg 69120, Germany. [12]Data Management Facility, German Cancer Research Center (DKFZ), Im Neuenheimer Feld 280, Heidelberg 69120, Germany. [13]Genomics Core Facility, European Molecular Biology Laboratory (EMBL), Meyerhofstrasse 1, Heidelberg 69117, Germany. [14]Department of Oncogenomics, AMC, University of Amsterdam, Meibergdreef 9, Amsterdam 1105 AZ, Netherlands. [15]The Arthur and Sonia Labbatt Brain Tumor Research Centre, Hospital for Sick Children, 555 University Avenue, Toronto, Ontario M5G 1X8, Canada. [16]Cell Networks Cluster of Excellence, University of Heidelberg, Heidelberg 69120, Germany. [17]Department of Advanced Diagnostic Technologies, IRCCS Azienda Ospedaliera Universitaria San Martino - IST Istituto Nazionale per la Ricerca sul Cancro, Lgo R. Benzi,10, Genoa 16132, Italy. [18]Center for Neuropathology and Prion Research, University of Munich, Feodor-Lynen-Strasse 23, Munich 81377, Germany. [19]Institute for Neuropathology, Evangelisches Krankenhaus, Remterweg 2, Bielefeld 33617, Germany. [20]Department of Paediatric Oncology and Haematology, Saarland University Hospital, Homburg 66421, Germany. [21]Institute for Pathology, Saarland University Hospital, Kirrberger Strasse, Homburg 66424, Germany. [22]Department of Neuropathology, Institute of Pathology, Würzburg University Josef-Schneider Strasse 2, Würzburg 97080, Germany. [23]Department of Neurosurgery, Heidelberg University Hospital, Im Neuenheimer Feld 400, Heidelberg 69120, Germany. [24]Clinical Cooperation Unit Pediatric Oncology, German Cancer Research Center (DKFZ), Im Neuenheimer Feld 280, Heidelberg 69120, Germany. [25]Department of Neuropathology, University of Heidelberg, Im Neuenheimer Feld 220, Heidelberg 69120, Germany. [26]Clinical Cooperation Unit Neuropathology, German Cancer Research Center (DKFZ), Im Neuenheimer Feld 220-221, Heidelberg 69120, Germany. [27]Department of Pediatric Oncology, Hematology & Immunology, Klinikum Stuttgart Olgahospital, Bismarckstrasse 8, Stuttgart 70176, Germany. [28]Department of Paediatric Haematology and Oncology, University Hospital Freiburg, Mathildenstrasse 1, Freiburg 79106, Germany. [29]Department of Hematology and Oncology, Children's University Hospital, Hoppe-Seyler Strasse 1, Tübingen 72076, Germany. [30]Department of Neurosurgery, University Hospital, Hoppe-Seyler Strasse 3, Tübingen 72076, Germany. [31]Children's Hospital Augsburg, Stenglinstrasse 2, Augsburg 86156, Germany. [32]Institute of Neuropathology, University Hospital Münster, Albert-Schweitzer-Campus 1, Münster 48149, Germany. [33]Departments of Pediatrics and Human Genetics, McGill University and the McGill University Health Center Research Institute, Montreal, Quebec H3Z 2Z3, Canada. [34]Department of Paediatric Haematology and Oncology, University Medical Center Hamburg-Eppendorf, Martinistrasse 52, Hamburg 20246, Germany. [35]Northern Institute for Cancer Research, Newcastle University, Royal Victoria Infirmary, Newcastle-upon-Tyne, NE1 4LP, UK. [36]School of Cancer and Enabling Sciences, University of Manchester, Manchester Academic Health Science Centre, Manchester, M13 9PL, UK. [37]Division of Molecular Histopathology, Department of Pathology, University of Cambridge, Cambridge CB2 0QQ, UK. [38]Division of Molecular Genome Analysis, German Cancer Research Center (DKFZ), Im Neuenheimer Feld 280, Heidelberg 69120, Germany. [39]Cnopf'sche Kinderklinik, Nürnberg Children's Hospital, St-Johannis-Mühlgasse 19, Nürnberg 90419, Germany. [40]Department of Neuropathology, Heinrich-Heine-University Düsseldorf, Moorenstrasse 5, Düsseldorf 40225, Germany. [41]Division of Neurosurgery and The Arthur and Sonia Labatt Brain Tumour Research Centre, Hospital for Sick Children, 555 University Avenue, Toronto, Ontario M5G 1X8, Canada. [42]Dana Farber Cancer Institute, 450 Brookline Avenue, Boston, Massachusetts 02215, USA. [43]Children's Hospital Boston, 300 Longwood Avenue, Boston, Massachusetts 02115, USA. [44]Institute of Pharmacy and Molecular Biotechnology, University of Heidelberg, Heidelberg 69120, Germany. [45]Bioquant Center, University of Heidelberg, Im Neuenheimer Feld 267, Heidelberg 69120, Germany.
*These authors contributed equally to this work.

LETTER

doi:10.1038/nature11329

Medulloblastoma exome sequencing uncovers subtype-specific somatic mutations

Trevor J. Pugh[1,2,3], Shyamal Dilhan Weeraratne[3,4], Tenley C. Archer[3,4], Daniel A. Pomeranz Krummel[5], Daniel Auclair[1], James Bochicchio[1], Mauricio O. Carneiro[1], Scott L. Carter[1], Kristian Cibulskis[1], Rachel L. Erlich[1], Heidi Greulich[1,2,3], Michael S. Lawrence[1], Niall J. Lennon[1], Aaron McKenna[1], James Meldrim[1], Alex H. Ramos[1,2,3], Michael G. Ross[1], Carsten Russ[1], Erica Shefler[1], Andrey Sivachenko[1], Brian Sogoloff[1], Petar Stojanov[1], Pablo Tamayo[1], Jill P. Mesirov[1], Vladimir Amani[3,4], Natalia Teider[3,4], Soma Sengupta[3,4], Jessica Pierre Francois[3,4], Paul A. Northcott[6], Michael D. Taylor[6], Furong Yu[7], Gerald R. Crabtree[7,8], Amanda G. Kautzman[7], Stacey B. Gabriel[1], Gad Getz[1], Natalie Jäger[9], David T. W. Jones[9], Peter Lichter[9], Stefan M. Pfister[9], Thomas M. Roberts[2,3], Matthew Meyerson[1,2,3,10], Scott L. Pomeroy[1,3,4] & Yoon-Jae Cho[1,3,4,7]

Medulloblastomas are the most common malignant brain tumours in children[1]. Identifying and understanding the genetic events that drive these tumours is critical for the development of more effective diagnostic, prognostic and therapeutic strategies. Recently, our group and others described distinct molecular subtypes of medulloblastoma on the basis of transcriptional and copy number profiles[2–5]. Here we use whole-exome hybrid capture and deep sequencing to identify somatic mutations across the coding regions of 92 primary medulloblastoma/normal pairs. Overall, medulloblastomas have low mutation rates consistent with other paediatric tumours, with a median of 0.35 non-silent mutations per megabase. We identified twelve genes mutated at statistically significant frequencies, including previously known mutated genes in medulloblastoma such as *CTNNB1*, *PTCH1*, *MLL2*, *SMARCA4* and *TP53*. Recurrent somatic mutations were newly identified in an RNA helicase gene, *DDX3X*, often concurrent with *CTNNB1* mutations, and in the nuclear co-repressor (N-CoR) complex genes *GPS2*, *BCOR* and *LDB1*. We show that mutant DDX3X potentiates transactivation of a TCF promoter and enhances cell viability in combination with mutant, but not wild-type, β-catenin. Together, our study reveals the alteration of WNT, hedgehog, histone methyltransferase and now N-CoR pathways across medulloblastomas and within specific subtypes of this disease, and nominates the RNA helicase DDX3X as a component of pathogenic β-catenin signalling in medulloblastoma.

Medulloblastomas are aggressive tumours of primitive neuroectodermal origin. More than one third of patients diagnosed with medulloblastoma succumb to their disease within 5 years[6] and surviving patients often have significant long-term adverse effects from current therapies. Identifying the underlying genetic events responsible for medulloblastomas can help guide the development of more effective therapies and refine the selection of currently available chemotherapy and radiotherapy. Recent efforts profiling transcriptional and DNA copy number changes in medulloblastoma have provided insights into the biological processes involved in these tumours and have underscored the molecular heterogeneity of this disease[2–4]. Based on these data, four broad subgroups have been established, known according to a consensus nomenclature as SHH, WNT, Group 3 and Group 4 (ref. 5).

The first genome-scale sequencing of protein coding regions in medulloblastoma was reported recently[7]. Altered genes encoding for histone modification proteins were identified in 20% of cases, most notably *MLL2* and *MLL3* (ref. 7). This initial survey was limited by a small discovery sample size (22 patients), lack of subtype-specific analysis, and use of Sanger sequencing technology insensitive to variants present at low allelic fraction. Here we survey coding somatic mutations at deeper coverage in a larger cohort of 92 medulloblastoma/normal pairs and assess these mutations in the context of specific molecular subtypes (Supplementary Table 1).

In total, 1,908 mutations were detected within 1,671 out of 18,863 genes sequenced to a median of 106× coverage (Supplementary Table 2). Confirmation of 20 candidate mutations in selected genes (*CTNNB1*, *DDX3X*, *SMARCA4*, *TP53* and *CTDNEP1*) was performed by amplification of 48 exons using a microfluidic PCR device (Fluidigm) followed by single-molecule real-time sequencing (SMRT, Pacific Biosciences) (Supplementary Information). Sequence data was unavailable for one *DDX3X* mutation because of poor PCR amplification from the sample. All remaining 19 mutations were confirmed by this orthogonal method (median 73 redundant sub-reads, range 3–287, Supplementary Fig. 1).

A median of 16 somatic mutations (12 non-silent, 4 silent) per tumour was identified, corresponding to a mutation rate of 0.35 non-silent mutations per megabase of callable sequence, less than most adult solid tumours and consistent with results from ref. 7. Six of the twelve most frequently mutated tumours were from the oldest patients (16–31 years at diagnosis), consistent with increased mutation frequency in adult versus childhood medulloblastomas ($P = 7.7 \times 10^{-5}$, Wilcoxon rank-sum test, Supplementary Fig. 2).

To identify genes mutated at statistically significant frequencies across our cohort, we used the MutSig algorithm[8] which takes into account gene size, sample-specific mutation rate, non-silent to silent mutation ratios, clustering within genes, and base conservation across species. In our cohort of 92 samples, we identified 12 significantly mutated genes ($q < 0.1$, Table 1 and Supplementary Table 3). Strikingly, these genes were not mutated in c5 (Group 3) and c4 (Group 4) tumours with extensive somatic copy number alteration (Fig. 1), indicating that these subtypes are driven primarily by structural variation, rather than base mutation. Not unexpectedly, *CTNNB1* (β-catenin) and *PTCH1* were the two most significantly mutated genes (see Table 1 and Fig. 1). Point mutations of *CTNNB1* in combination with loss of chromosome 6 were found in all WNT subgroup tumours and were concurrent with several other recurrently mutated genes, namely *CSNK2B*, *DDX3X*, *TP53* and *SMARCA4*. Mutations involving *PTCH1* occurred exclusively in SHH subgroup tumours and mutations of genes associated with the hedgehog pathway were also restricted to this subgroup ($P < 0.0001$, Fisher's exact test). All but one of the tumours with *PTCH1* mutations had somatic loss of 9q, resulting in hemizygosity for the mutant allele. The remaining tumour had apparent copy neutral loss-of-heterozygosity of 9q22. Other

[1]Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA. [2]Center for Cancer Genome Discovery, Departments of Medical Oncology and of Biological Chemistry and Molecular Pharmacology, Dana-Farber Cancer Institute, Boston, Massachusetts 02115, USA. [3]Harvard Medical School, Boston, Massachusetts 02115, USA. [4]Department of Neurology, Children's Hospital Boston, Boston, Massachusetts 02115, USA. [5]Brandeis University, Waltham, Massachusetts 02453, USA. [6]The Hospital for Sick Children, Toronto, Ontario M5G 1X8, Canada. [7]Departments of Neurology and Neurosurgery, Stanford University School of Medicine, Stanford, California 94305, USA. [8]Howard Hughes Medical Institute at Stanford University, Stanford, California 94305, USA. [9]German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany. [10]Department of Pathology, Brigham and Women's Hospital, Boston, Massachusetts 02115, USA.

Table 1 | Genes mutated at a statistically significant frequency in 92 medulloblastomas.

Gene	Description	Mutations	Patients	Unique sites	Silent	Missense	Indel or null	Double null	q
CTNNB1	β-catenin	6	6	4	0	6	0	0	$<1.8 \times 10^{-11}$
PTCH1	Patched homologue 1 (*Drosophila*)	7	7	7	0	0	7	0	4.0×10^{-9}
MLL2	Myeloid/lymphoid or mixed-lineage leukaemia 2	10	8	10	0	2	4	4	4.0×10^{-9}
DDX3X	DEAD box polypeptide 3, X-linked	7	7	7	0	7	0	0	2.3×10^{-8}
GPS2	G protein pathway suppressor 2	3	3	3	0	1	2	0	1.2×10^{-4}
TP53	Tumour protein p53	3	3	3	0	3	0	0	0.039
KDM6A	UTX, lysine (K)-specific demethylase 6A	3	3	3	0	2	1	0	0.042
BCOR	BCL6 co-repressor	3	3	3	0	0	3	0	0.046
SMARCA4	ATP-dependent helicase	4	4	3	0	4	0	0	0.046
LDB1	LIM domain binding 1	2	2	2	0	1	1	0	0.047
CTDNEP1	CTD nuclear envelope phosphatase 1	2	2	2	0	0	2	0	0.047
CSNK2B	Casein kinase 2, β polypeptide	2	2	2	0	2	0	0	0.071

Null, nonsense, frameshift or splice-site mutations; double null, null mutations co-occurring in a single tumour; q, q-value, false discovery rate (Benjamini–Hochberg procedure). See Supplementary Table 3 for further statistics and subtype analysis.

somatic mutations of hedgehog pathway members include a splice site mutation in *SUFU*, an in-frame deletion in *WNT6*, and missense mutations in *GLI2*, *SMO*, *PRKACA*, *WNT2* and *WNT2B*.

Two patients with SHH subgroup tumours had germline variants in *PTCH1*, one with somatic loss of 9q resulting in hemizygosity for a loss-of-function germline allele (MD-085, c.3030delC, p.Asn1011Thrfs*38), and the other with a substitution previously reported in patients with holoprosencephaly (MD-286, p.T1052M, ref. 9). Two additional cases (MD-097 and MD-335) had loss-of-function variants in *SUFU* (1 frameshift deletion and 1 nonsense) that began as heterozygotes in the germline and became hemizygous in the tumour, due to somatic loss of chromosome 10 in one case and copy neutral loss-of-heterozygosity in the other.

MLL2 was also subject to recurrent inactivating mutations, consistent with findings from ref. 7 and providing further evidence for dysregulated histone modification in medulloblastoma. Indeed, six

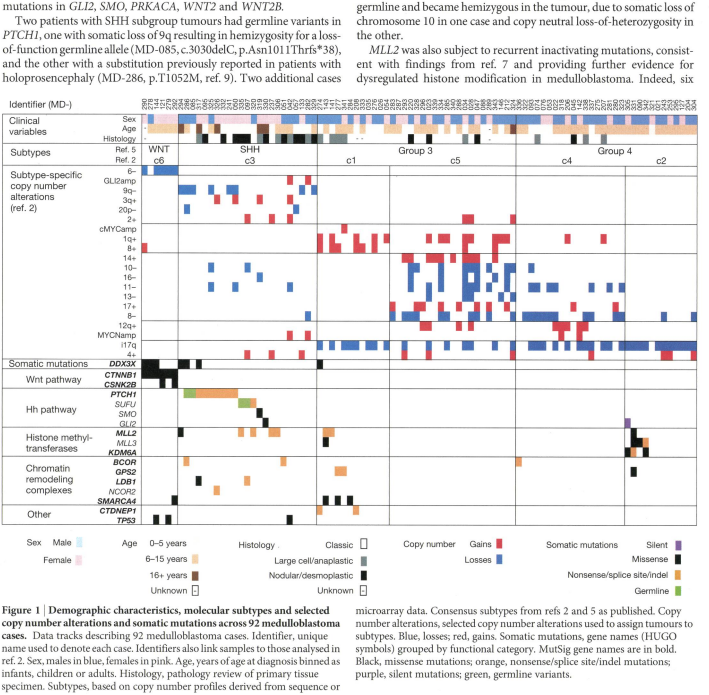


Figure 1 | Demographic characteristics, molecular subtypes and selected copy number alterations and somatic mutations across 92 medulloblastoma cases. Data tracks describing 92 medulloblastoma cases. Identifier, unique name used to denote each case. Identifiers also link samples to those analysed in ref. 2. Sex, males in blue, females in pink. Age, years of age at diagnosis binned as infants, children or adults. Histology, pathology review of primary tissue specimen. Subtypes, based on copy number profiles derived from sequence or microarray data. Consensus subtypes from refs 2 and 5 as published. Copy number alterations, selected copy number alterations used to assign tumours to subtypes. Blue, losses; red, gains. Somatic mutations, gene names (HUGO symbols) grouped by functional category. MutSig gene names are in bold. Black, missense mutations; orange, nonsense/splice site/indel mutations; purple, silent mutations; green, germline variants.

of the twelve most significantly mutated genes are involved in histone modification and/or related chromatin remodelling complexes (*MLL2*, *GPS2*, *KDM6A*, *BCOR*, *SMARCA4* and *LDB1*; see Table 1). As a gene set, histone methyltransferases (HMTs) were enriched for somatic mutation with 21 tumours having apparent, predominantly loss-of-function, HMT mutations ($q = 5.8 \times 10^{-9}$; Fig. 2 and Supplementary Table 4).

Subtype-specific MutSig analysis identified additional significant mutations of histone-modifying genes, *MLL3* and *HDAC2*, in Group 4 tumours along with *KDM6A* mutations ($q = 0.039$ and 0.066, see Supplementary Table 3). Mutations in KDM6A, interestingly, occurred exclusively in tumours with an i17q as the sole autosomal alteration ($P = 0.0023$, Fisher's exact test) with the one female case with *KDM6A* mutation also having loss of a chromosome X. Notably, the two 'i17q only' tumours without *KDM6A* mutations had other histone-modifying enzymes mutated, namely *THUMPD3*, *ZMYM3* and *MLL3*, perhaps suggesting a distinct biology for tumours with this karyotype.

Mutations in several genes encoding components of the nuclear co-repressor (N-CoR) complex were observed at a statistically significant frequency: *BCOR* in 3 tumours, *GPS2* in 3 tumours, and *LDB1* in 2 tumours. *BCOR* mutations have recently been reported at high frequency in retinoblastoma[10] and in 'copy-neutral' acute myelogenous leukaemia[11]. *BCOR* is located on the X-chromosome and two hemizygous frameshift mutants were found in tumours from males (allele fractions 0.90 and 0.92). A third nonsense mutation was also found in a male but at low allelic fraction (0.12), indicating a subclonal event. Two out of three *BCOR* mutations occurred in SHH subgroup tumours. *LDB1* missense and nonsense mutations were found in two additional SHH tumours, both appearing hemizygous due to loss of 10q and complete chromosome 10 loss, respectively (allele fractions 0.81 and 0.78). Both *BCOR* and *LDB1* promote assembly of the repressive N-CoR complex[12] and harbour apparent loss of function mutations. *GPS2*, which encodes a critical subunit of the N-CoR complex, a repressor of JNK/MAPK signalling through partnership with histone deacetylases[12], was mutated in two Group 3 tumours. The *GPS2* mutations cluster within amino acids 53–90, the domain critical for heterodimerization with *NCOR2* (also known as SMRT) and interacting with a TBL1 amino-terminal domain tetramer to assemble the N-CoR repression complex[12]. Finally, an additional nonsense mutation in *NCOR2* was identified in a single SHH subgroup tumour, underscoring the central role of N-CoR dysregulation in medulloblastoma development and particularly within the SHH subgroup.

Several genes encoding subunits of the SWI/SNF-like chromatin-remodelling complex were also mutated in our cohort, including significant recurrent mutations of *SMARCA4(Brg/BAF190)*, which encodes a DNA helicase with ATPase activity[13] and has been reported to be mutated in lung, ovarian, and pancreatic cancers[14] as well as medulloblastoma[15]. In our cohort, *SMARCA4(Brg/BAF190)* mutations clustered in helicase domains and occurred in three Group 3 tumours (significant within the c1 subtype, $q = 0.019$), and one WNT tumour. In addition, mutations were found in the alternative ATPase subunit *SMARCD2(Brm)* (missense at a highly conserved residue) and two other members of the SWI/SNF complex, *ARID1B(BAF250b)* (2 base pairs (bp) frameshift deletion) and *SMARCC2(BAF170)* (splice site). These were all apparent loss-of-function mutations and occurred in SHH tumours. Thus, it seems that disruption of this complex is frequent across medulloblastomas.

New and hemizygous mutations were found in *CTDNEP1* (previously known as *DULLARD*), a phosphatase with roles in *Xenopus* neural development through regulation of BMP receptors[16], and as a direct regulator of LIPIN, an integral component of the mTOR complex[17]. *CTDNEP1* mutations were found in two Group 3 tumours (significant within the subtype, $q = 0.0087$), a 2-bp frameshift deletion and a substitution disruptive of a splice site. Both tumours have i17q chromosomes, resulting in loss of the wild-type allele at 17p13.

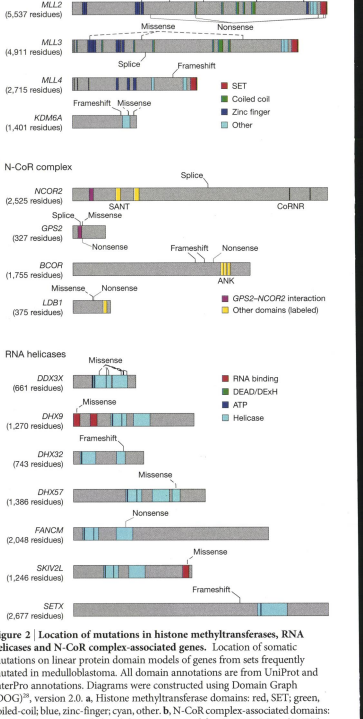


Figure 2 | Location of mutations in histone methyltransferases, RNA helicases and N-CoR complex-associated genes. Location of somatic mutations on linear protein domain models of genes from sets frequently mutated in medulloblastoma. All domain annotations are from UniProt and InterPro annotations. Diagrams were constructed using Domain Graph (DOG)[38], version 2.0. **a**, Histone methyltransferase domains: red, SET; green, coiled-coil; blue, zinc-finger; cyan, other. **b**, N-CoR complex-associated domains: purple, anti-parallel coiled-coil domains required for GPS2–NCOR2 (SMRT) interaction[12]; yellow, other interaction domains as labelled (SANT domains binds DNA, CoRNR domains binds nuclear receptors, ANK repeats mediate a diversity of protein–protein interactions, and LIM-binding domains bind a common protein structural motif). **c**, RNA helicase domains: cyan, helicase and helicase-associated (InterPro); red, RNA-binding and RNA polymerase sigma factor (InterPro); blue, ATP binding site; green, DEAD or DExH box motif. See Supplementary Table 1 for UniProt protein model identifiers.

Mutations in *DDX3X*, an ATP-dependent RNA helicase with functions in transcription, splicing, RNA transport and translation[18], were found in seven tumours, including half of the WNT pathway tumours ($P = 0.005$, Fisher's exact test) and several SHH subgroup tumours.

DDX3X mutations have recently been reported at low frequency in five other tumour types (Catalogue of Somatic Mutations in Cancer, COSMIC[15]) but the significance of these mutations for DDX3X function remains unclear. To understand the consequence of observed point mutations on the physical structure of DDX3X, we mapped the mutations onto the previously reported crystal structure of DDX3X[19] and its orthologue DDX4 (also known as VASA, ref. 20) (Fig. 3a; Supplementary Fig. 3 and Supplementary Table 5). The mutations seem to cluster in two structural domains, a helicase ATP-binding domain (residues 211–403) and a helicase carboxy-terminal domain (residues 414–575). The location of these mutations indicates that they may alter DDX3X–RNA interaction (Fig. 3a and Supplementary Table 5).

As half of the β-catenin mutated tumours contained concurrent *DDX3X* mutations, we investigated whether DDX3X could enhance the of ability β-catenin to transactivate a TCF4-luciferase reporter (TOPflash) and if DDX3X/β-catenin co-expression had a measureable effect on cell viability/proliferation. In combination with wild-type β-catenin, neither wild-type nor mutant DDX3X alone significantly transactivated the TOPflash reporter. However, in combination with mutant β-catenin (S33Y substitution), the majority of DDX3X point mutants in our cohort potentiated reporter activity ($P < 0.05$, Fig. 3b). This potentiation was also apparent in cell viability assays in both HeLa (data not shown) and D425 medulloblastoma cell lines ($P < 0.05$, Fig. 3c).

Given the apparent importance of *DDX3X* mutations in medulloblastoma, we searched the genes listed in the RNA Helicase Database (http://www.rnahelicase.org/) for low frequency mutations in medulloblastoma. We found five tumours with mutations within RNA helicase or RNA binding domains of *DHX9, DHX32, DHX57, FANCM* and *SKIV2L* (Fig. 2 and Supplementary Table 6). The missense mutations were at conserved residues and predicted to be deleterious by the software packages SIFT, AlignGVGD and PolyPhen2. In addition, a frameshift insertion in *SETX* occurs upstream of, and probably disrupts, its RNA helicase domain. Overall, 15% of medulloblastomas seem to have some disruption of RNA helicase activity.

In summary, we report a next-generation sequencing analysis of medulloblastoma, the most common malignant brain tumour in children. Our results reveal mutations in several known pathways such as histone methylation (*MLL2* and others), sonic hedgehog (*PTCH1*, *SUFU* and others) and Wnt (*CTNNB1* and others), and also previously unrecorded mutations in genes including *DDX3X*, *BCOR*, *LDB1* and *GPS2*. Our preliminary functional studies implicate DDX3X as a candidate component of pathogenic WNT/β-catenin signalling. In a broader sense, *DDX3X* mutations have recently been reported in chronic lymphocytic leukaemia[21] and head and neck cancers[22], both of which have subsets of tumours with dysregulated WNT signalling. Studies investigating whether mutant DDX3X functions together with β-catenin in these contexts should provide additional insights into this multifaceted molecule and open potential avenues for novel therapies. Finally, the delineation of nuclear receptor co-repressor complex molecules as altered in medulloblastoma provides new insight into the pathogenesis of this deadly childhood disease.

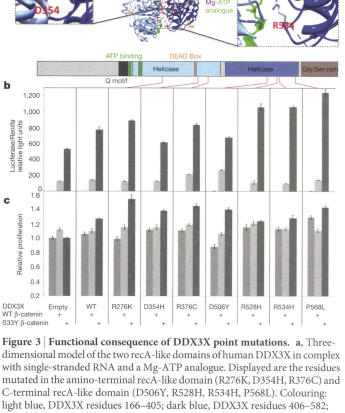


Figure 3 | Functional consequence of DDX3X point mutations. a, Three-dimensional model of the two recA-like domains of human DDX3X in complex with single-stranded RNA and a Mg-ATP analogue. Displayed are the residues mutated in the amino-terminal recA-like domain (R276K, D354H, R376C) and C-terminal recA-like domain (D506Y, R528H, R534H, P568L). Colouring: light blue, DDX3X residues 166–405; dark blue, DDX3X residues 406–582; cyan, single-stranded RNA; magenta and green, Mg-ATP analogue. Molecular graphics images were produced using the University of San Francisco Chimera package[29] (http://www.cgl.ucsf.edu/chimera). **b**, Mutant DDX3X potentiates mutant β-catenin transactivation of TOPflash promoter. Represented is relative luciferase activity in 293T cells co-transfected with TOPflash reporter, FOPflash control, and either wild-type or mutant DDX3X in combination with wild-type or mutant β-catenin. One-dimensional model of DDX3X displayed above bar graphs to illustrates the position of the mutations. WT, wild type. **c**, Cell viability assays of medulloblastoma D425 cells stably transduced with either wild-type or mutant DDX3X lentivirus in combination with either wild-type or mutant β-catenin lentivirus. For **b** and **c**, error bars depict the standard deviation of the mean from five replicate experiments performed for each condition. Student's *t*-tests were performed to evaluate significance of differences in TOPflash intensity or cell proliferation value distributions as follows: increases with DDX3X alone versus empty vector, increases with wild-type β-catenin versus DDX3X alone, increases with mutant β-catenin versus DDX3X alone, and increases with mutant β-catenin versus wild-type β-catenin.

METHODS SUMMARY

Informed consent was provided by families of medulloblastoma patients treated at Children's Hospital Boston, The Hospital for Sick Children, Toronto, Canada, and institutions contributing to the Children's Oncology Group/Cooperative Human Tissue Network, under approval and oversight by their respective Internal Review Boards. All tumours were obtained at the initial surgical resection and recurrent tumours were excluded from our analysis. Haematoxylin- and eosin-stained slides of tumour samples were reviewed by a pathologist to confirm the diagnosis of medulloblastoma, determine histological subtype when possible, and assess tumour purity. DNA was isolated from tumour specimens and matched peripheral blood as previously described[2]. Exome sequencing of DNA from 92 tumour/normal pairs was performed using in-solution hybrid-capture of 193,094 exons from 18,863 micro RNA (miRNA)- and protein-coding genes, followed by sequencing of 76 bp paired-end reads using Illumina sequencing-by-synthesis technology[23]. Reads were aligned to human genome build GRCh37[24] using a Burrows–Wheeler aligner (BWA)[25]. The ~33-megabase target region was sequenced to 106× mean coverage in each sample (range 73–234). Gene expression data and copy number profiles (derived from SNP microarrays or sequence data) were used to assign each tumour to a subgroup using published criteria[2]. Our cohort consisted of 6 WNT (c6), 23 SHH (c3), 33 Group 3 (12 c1, 21 c5), and 30 Group 4 (12 c2, 18 c4) tumours (see Supplementary Table 1 for case annotations). Mutations were detected using muTect, annotated using Oncotator[26], and manually reviewed using the Integrated Genomics Viewer (IGV)[27]. For validation, PCR on Access Array microfluidic chips (Fluidigm) was followed by single-molecule real-time sequencing (Pacific Biosciences) as per manufacturer's instructions. Sub-reads were extracted and assigned to samples using manufacturer's and custom software, and aligned to the hg19 (GRCh37) build of the human reference genome sequence using BWA-SW[25]. Candidate mutations were confirmed by manual review using IGV[27] (Supplementary Fig. 1). See Supplementary Information and http://www.broadinstitute.org/cancer/cga/ for complete descriptions of materials and methods.

Received 3 February; accepted 15 June 2012.
Published online 22 July 2012.

1. Central Brain Tumor Registry of the United States. *Statistical Report: Primary Brain and Central Nervous System Tumors Diagnosed in the United States in 2004–2007* http://www.cbtrus.org/2011-NPCR-SEER/WEB-0407-Report-3-3-2011.pdf (CBTRUS, 2011).
2. Cho, Y.-J. *et al.* Integrative genomic analysis of medulloblastoma identifies a molecular subgroup that drives poor clinical outcome. *J. Clin. Oncol.* **29,** 1424–1430 (2011).
3. Kool, M. *et al.* Integrated genomics identifies five medulloblastoma subtypes with distinct genetic profiles, pathway signatures and clinicopathological features. *PLoS ONE* **3,** e3088 (2008).
4. Remke, M. *et al.* Adult medulloblastoma comprises three major molecular variants. *J. Clin. Oncol.* **29,** 2717–2723 (2011).
5. Taylor, M. D. e. t. a. l. Molecular subgroups of medulloblastoma: the current consensus. *Acta Neuropathol.* **123,** 465–472 (2012).
6. Smoll, N. R. Relative survival of childhood and adult medulloblastomas and primitive neuroectodermal tumors (PNETs). *Cancer* **118,** 1313–1322 (2012).
7. Parsons, D. W. *et al.* The genetic landscape of the childhood cancer medulloblastoma. *Science* **331,** 435–439 (2011).
8. Getz, G. *et al.* Comment on "The consensus coding sequences of human breast and colorectal cancers.". *Science* **317,** 1500 (2007).
9. Ming, J. E. *et al.* Mutations in PATCHED-1, the receptor for SONIC HEDGEHOG, are associated with holoprosencephaly. *Hum. Genet.* **110,** 297–301 (2002).
10. Zhang, J. *et al.* A novel retinoblastoma therapy from genomic and epigenetic analyses. *Nature* **481,** 329–334 (2012).
11. Grossmann, V. *et al.* Whole-exome sequencing identifies somatic mutations of *BCOR* in acute myeloid leukemia with normal karyotype. *Blood* **118,** 6153–6163 (2011).
12. Oberoi, J. *et al.* Structural basis for the assembly of the SMRT/NCoR core transcriptional repression machinery. *Nature Struct. Mol. Biol.* **18,** 177–184 (2011).
13. Baek, S. H. *et al.* Regulated subset of G1 growth-control genes in response to derepression by the Wnt pathway. *Proc. Natl Acad. Sci. USA* **100,** 3245–3250 (2003).
14. Wilson, B. G. & Roberts, C. W. M. SWI/SNF nucleosome remodellers and cancer. *Nature Rev. Cancer* **11,** 481–492 (2011).
15. Futreal, P. A. *et al.* A census of human cancer genes. *Nature Rev. Cancer* **4,** 177–183 (2004).
16. Satow, R., Kurisaki, A., Chan, T.-c., Hamazaki, T. S. & Asashima, M. Dullard promotes degradation and dephosphorylation of BMP receptors and is required for neural induction. *Dev. Cell* **11,** 763–774 (2006).
17. Peterson, T. R. *et al.* mTOR complex 1 regulates lipin 1 localization to control the SREBP pathway. *Cell* **146,** 408–420 (2011).
18. Garbelli, A., Beermann, S., Di Cicco, G., Dietrich, U. & Maga, G. A motif unique to the human dead-box protein DDX3 is important for nucleic acid binding, ATP hydrolysis, RNA/DNA unwinding and HIV-1 replication. *PLoS ONE* **6,** e19810 (2011).
19. Högbom, M. *et al.* Crystal structure of conserved domains 1 and 2 of the human DEAD-box helicase DDX3X in complex with the mononucleotide AMP. *J. Mol. Biol.* **372,** 150–159 (2007).
20. Sengoku, T., Nureki, O., Nakamura, A., Kobayashi, S. & Yokoyama, S. Structural basis for RNA unwinding by the DEAD-box protein *Drosophila* Vasa. *Cell* **125,** 287–300 (2006).
21. Wang, L. *et al. SF3B1* and other novel cancer genes in chronic lymphocytic leukemia. *N. Engl. J. Med.* **365,** 2497–2506 (2011).
22. Stransky, N. *et al.* The mutational landscape of head and neck squamous cell carcinoma. *Science* **333,** 1157–1160 (2011).
23. Bentley, D. R. *et al.* Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* **456,** 53–59 (2008).
24. Genome Reference Consortium. Human Genome Overview http://www.ncbi.nlm.nih.gov/projects/genome/assembly/grc/human/ (2012).
25. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25,** 1754–1760 (2009).
26. Ramos, A. *et al.* Oncotator http://www.broadinstitute.org/oncotator/ (2012).
27. Robinson, J. T. *et al.* Integrative genomics viewer. *Nature Biotechnol.* **29,** 24–26 (2011).
28. Ren, J. *et al.* DOG 1.0: illustrator of protein domain structures. *Cell Res.* **19,** 271–273 (2009).
29. Pettersen, E. F. *et al.* UCSF Chimera–a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25,** 1605–1612 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by NIH grants NHGRI U54HG003067 to E. S. Lander (E.S., D.A., S.B.G., G.G., M.M.); R01CA109467 (S.L.P., J.P.M.); R01CA105607 (H.G., T.M.R., M.M., S.L.P.); P30 HD18655 (S.L.P.); R01 CA030002 and CA050661 (T.M.R.); R01 NS046789 (G.R.C.); R01 CA154480 (P.T.); R25NS070682 (S.S.) and R01CA148699 (M.D.T.); St. Baldrick's Foundation Scholar Award and the Beirne Faculty Scholar endowment and Center for Children's Brain Tumors at Stanford University (Y.-J.C.); German Cancer Aid (109252) and the BMBF ICGC-PedBrain project (N.J., D.T.W.J., P.L., S.M.P.); HHMI (G.R.C.); the Pediatric Brain Tumor Foundation (M.D.T.); Canadian Institutes of Health Research Fellowship (T.J.P.); Restracomp funding from the Hospital for Sick Children (P.A.N.); and the Mullarkey Research Fund (S.L.P.). We thank Children's Oncology Group and the Cooperative Human Tissue Network for providing tumour samples, the staff of the Broad Institute Biological Samples, Genome Sequencing and Genetic Analysis Platforms for their assistance in genomic processing of samples and generating the sequencing data used in this analysis, K. Keho and M. Brown at Pacific Biosciences for technical support with sample barcoding methods, and L. Gaffney of Broad Institute Communications for assistance with figure layout and design.

Author Contributions Y.-J.C., M.M. and S.L.P. conceived the project. Y.-J.C., T.J.P., M.M. and S.L.P. wrote the manuscript with input from co-authors. S.D.W., T.C.A., J.P.F., S.S., N.T., Y.-J.C., A.G.K. and F.Y. performed functional characterization studies. D.A.P.K. generated *in silico* structural modelling of DDX3X mutations. T.J.P. conducted the bioinformatic analysis, supported by S.L.C., P.S., K.C., M.S.L., A.M., A.H.R., A.S., H.G., P.T., J.P.M., N.J. and D.T.W.J.; D.A., E.S., S.B.G., and G.G. facilitated transfer, sequencing and analysis of samples. P.A.N. and M.D.T. provided tissues for analysis. Y.-J.C., J.P.F. and V.A. processed tumour and blood samples for study. G.R.C. generated reagents used in functional characterization studies. P.L., S.M.P. and T.M.R. assisted with interpretation of results. J.B., M.O.C., R.L.E., N.J.L., J.M., M.G.R., C.R. and B.S. performed microfluidic PCR and single-molecule real-time sequencing for validation analysis.

Author Information Sequence data used for this analysis are available in dbGaP under accession phs000504.v1.p1. Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the full-text HTML version of the paper at www.nature.com/nature. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to Y.-J.C. (yjcho1@stanford.edu), M.M. (matthew_meyerson@dfci.harvard.edu) or S.L.P. (scott.pomeroy@childrens.harvard.edu).

LETTER

doi:10.1038/nature11362

Targeting nuclear RNA for *in vivo* correction of myotonic dystrophy

Thurman M. Wheeler[1,2], Andrew J. Leger[3], Sanjay K. Pandey[4], A. Robert MacLeod[4], Masayuki Nakamori[1,2], Seng H. Cheng[3], Bruce M. Wentworth[3], C. Frank Bennett[4] & Charles A. Thornton[1,2]

Antisense oligonucleotides (ASOs) hold promise for gene-specific knockdown in diseases that involve RNA or protein gain-of-function effects. In the hereditary degenerative disease myotonic dystrophy type 1 (DM1), transcripts from the mutant allele contain an expanded CUG repeat[1–3] and are retained in the nucleus[4,5]. The mutant RNA exerts a toxic gain-of-function effect[6], making it an appropriate target for therapeutic ASOs. However, despite improvements in ASO chemistry and design, systemic use of ASOs is limited because uptake in many tissues, including skeletal and cardiac muscle, is not sufficient to silence target messenger RNAs[7,8]. Here we show that nuclear-retained transcripts containing expanded CUG (CUG^{exp}) repeats are unusually sensitive to antisense silencing. In a transgenic mouse model of DM1, systemic administration of ASOs caused a rapid knockdown of CUG^{exp} RNA in skeletal muscle, correcting the physiological, histopathologic and transcriptomic features of the disease. The effect was sustained for up to 1 year after treatment was discontinued. Systemically administered ASOs were also effective for muscle knockdown of *Malat1*, a long non-coding RNA (lncRNA) that is retained in the nucleus[9]. These results provide a general strategy to correct RNA gain-of-function effects and to modulate the expression of expanded repeats, lncRNAs and other transcripts with prolonged nuclear residence.

Antisense silencing by the RNase H-dependent mechanism entails a three-step process of oligonucleotide hybridization to its cognate RNA, cleavage of the target by RNase H1 and exonuclease degradation of the cleavage fragments. The relative efficiency of this mechanism in the nucleus and cytoplasm is uncertain. Observations that ASOs shuttle from cytoplasm to nucleus[10] and that targeting intronic sequences[11] or nuclear RNA[12] can silence gene expression indicate that antisense is active in the nucleus. However, activity in the cytoplasm is less clear. Although RNase H1 is not restricted to the nucleus[13], recent studies indicate that the non-nuclear fraction is confined to mitochondria[14]. This suggests that ASO•RNase H cleavage is mainly a nuclear process, whose potency could be maximized by targeting transcripts with long nuclear residence.

To test this idea, we used a transgenic mouse model of DM1. HSA^{LR} transgenic mice express CUG^{exp} RNA at high levels in skeletal muscle. Human DM1 is caused by an expanded CTG repeat in the 3′ untranslated region (UTR) of dystrophia myotonica-protein kinase (*DMPK*)[3], whereas in HSA^{LR} mice the expanded repeat is in the 3′ UTR of a human skeletal actin (h*ACTA1*) transgene[6]. In both conditions the CUG^{exp} transcripts are retained in nuclear foci, along with splicing factors in the muscleblind-like (MBNL) protein family. Muscleblind sequestration leads to misregulated alternative splicing and other changes of the muscle transcriptome[15–17]. The RNA toxicity was mitigated in mice by CAG-repeat morpholino oligomers that compete with MBNL proteins for CUG^{exp} binding, without activating RNase H. However, this approach required direct injection into a single muscle, followed by *in vivo* electroporation, a method to load muscle fibres with oligomers[18]. As an alternative, RNase H-active ASOs could produce widespread correction, provided that uptake of circulating ASOs was sufficient to induce target cleavage.

We identified ASOs showing a strong knockdown of h*ACTA1* in tissue culture, good tolerability when systemically administered in wild-type mice, and activity against h*ACTA1*-CUG^{exp} transcripts *in vivo* when electroporated into muscle (Supplementary Figs 1–3). The ASOs had 2′-*O*-methoxyethyl (MOE) modifications at both ends to maximize biostability, and a central gap of 10 unmodified nucleotides to support RNase H activity (MOE gapmers; Supplementary Table 1). We tested three of the ASOs in HSA^{LR} transgenic mice by subcutaneous injection of 25 mg kg^{-1} twice weekly (Fig. 1a). After 4 weeks of administration (8 injections), ASO 445236 reduced the level of CUG^{exp} RNA in hindlimb muscles by more than 80% (Fig. 1b). Another ASO targeting the 3′ UTR, downstream of the repeat tract, also showed strong CUG^{exp} reduction, whereas an ASO targeting the 5′ UTR, or three oligonucleotides against other targets, had no effect (Fig. 1b, c).

RNase H cleavage of mRNA is usually followed by rapid decay of cleavage fragments. However, stable cleavage fragments are observed occasionally[19], and the CUG^{exp} tract forms extensive hairpins[20] and ribonucleoprotein complexes[21] that could inhibit exonuclease activity. The failure of antisense targeting in the 5′ UTR also raised the possibility that cleavage downstream of the repeat tract was required for efficient silencing. We therefore tested an additional ASO, 190401, targeting the h*ACTA1* coding region, and found that it also was highly effective (Fig. 1d). Furthermore, northern blot analysis using a CAG-repeat probe showed no evidence for a stable CUG^{exp} cleavage fragment (Fig. 1e), and *in situ* hybridization showed reduction of nuclear CUG^{exp} foci (Supplementary Fig. 4). These results indicate that expanded CUG repeats are degraded after a cleavage event 5′ or 3′ of the repeat tract.

Reduction of CUG^{exp} RNA would be expected to release sequestered MBNL1 protein and improve its splicing regulatory activity. Consistent with this prediction, alternative splicing of four MBNL1-dependent exons, *Serca1* (also known as *Atp2a1*) exon 22, titin (*Ttn*) exon 362, *Zasp* (also known as *Ldb3*) exon 11, and *Clcn1* chloride ion channel exon 7a, was normalized (Fig. 1f, g and Supplementary Figs 5 and 6)[15]. The *Clcn1* splicing defect causes loss of channel function, repetitive action potentials and delayed muscle relaxation (myotonia)[22], a cardinal feature of the disease. Blind analysis showed that myotonic discharges in hindlimb muscles were eliminated by the active ASOs (Fig. 1h), indicating rescue of *Clcn1* function.

In addition to splicing defects, expression of CUG^{exp} RNA or ablation of *Mbnl1* causes extensive remodelling of the muscle transcriptome[16,17,23]. We used microarrays to examine transcriptomic effects of ASOs. Principle component analysis showed that gene expression in ASO-treated HSA^{LR} mice was shifted towards wild-type mice, indicating an overall trend for transcriptome normalization (Fig. 2a). Among transcripts that were up- or downregulated in HSA^{LR} muscle, more

[1]Department of Neurology, University of Rochester, 601 Elmwood Avenue, Rochester, New York, 14642, USA. [2]Center for Neural Development and Disease, University of Rochester, 601 Elmwood Avenue, Rochester, New York, 14642, USA. [3]Genzyme Corporation, 49 New York Avenue, Framingham, Massachusetts 01701, USA. [4]Isis Pharmaceuticals, 2855 Gazelle Court, Carlsbad, California 92010, USA.

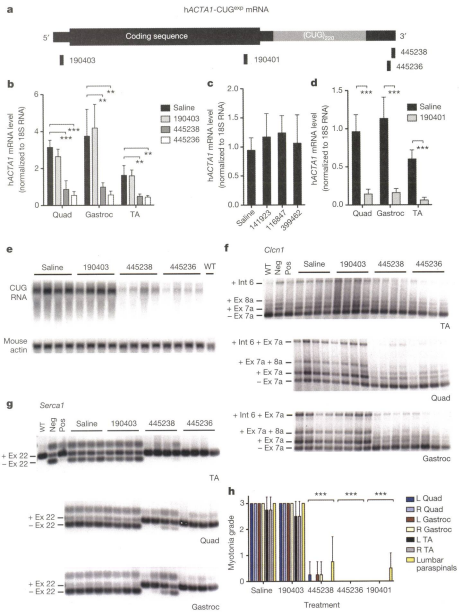


Figure 1 | Systemic administration of 2′-*O*-(2-methoxyethyl) ASOs in the *HSA*^{LR} transgenic mouse model of DM1. **a**, Location of ASO-targeting sequences relative to the h*ACTA1* coding region and the expanded CUG repeat in the 3′ UTR. **b**, Quantitative real-time RT–PCR of h*ACTA1*-CUG^{exp} mRNA in quadriceps (Quad), gastrocnemius (Gastroc) and tibialis anterior (TA) muscle in *HSA*^{LR} mice treated with the indicated ASOs by subcutaneous injection of 25 mg kg^{-1} twice weekly for 4 weeks. Muscle tissue was obtained 1 week after the final dose ($n = 4$ per group). The mean levels of transgene mRNA ± s.d. are shown. $^{**}P < 0.001$, $^{***}P < 0.0001$ (one-way analysis of variance (ANOVA)). **c**, h*ACTA1*-CUG^{exp} transcript levels in quadriceps are not affected by ASOs targeting unrelated transcripts (141923, randomer; 116847, *Pten*; 399462, *Malat1*; $n = 4$ per group; same dose as in **b**). Error bars are mean ± s.d. **d**, Knockdown of h*ACTA1*-CUG^{exp} mRNA in muscle by ASO 190401 ($n = 4$ per group; same dose as in **b**). Error bars are mean ± s.d. $^{***}P \leq 0.0005$ (*t*-test). **e**, Northern analysis of RNA from quadriceps muscle. CUG^{exp} RNA was detected using a $(CAG)_9$ oligonucleotide probe. Mouse actin serves as loading control. **f**, **g**, RT–PCR analysis of alternative splicing of *Clcn1* (**f**) and *Serca1* (**g**) transcripts. For *Clcn1*, only the –ex7a isoform encodes a functional ion channel. –ex7a, exon 7a exclusion; +ex7a, exon 7a inclusion; –ex22, exon 22 exclusion; +ex22, exon 22 inclusion; neg, negative control mice injected with GAC25 morpholino; pos, positive control mice injected with CAG25 morpholino; WT, FVB/N wild-type mice. **h**, Blind analysis of myotonia using EMG, 1 week after the final dose ($n = 4$ mice per group). Error bars are mean ± s.d. $^{***}P < 0.0001$ for ASO-treated versus saline-treated muscles (two-way ANOVA).

than 85% were normalized or partially corrected by ASOs, without evidence of off-target effects (Fig. 2b, and Supplementary Fig. 7 and Supplementary Table 2). These results confirm that ASOs caused an overall improvement of the muscle transcriptome.

ASO effects were evident within 2 weeks (Supplementary Fig. 8) and were dose-dependent. A threefold dose reduction (8.5 mg kg^{-1} biweekly for 4 weeks) caused partial myotonia and splicing correction, whereas a tenfold dose reduction (2.5 mg kg^{-1}) caused partial myotonia correction in tibialis anterior but not in quadriceps (Supplementary Fig. 9a–c), the latter muscle having higher basal levels of CUG^{exp} expression[18]. Serum chemistries showed no evidence for renal or liver toxicity (Supplementary Fig. 10).

A uniform finding in previous studies of MOE gapmer ASOs was that systemic administration failed to cause significant target reduction in muscle, despite efficient knockdown in liver ($n = 12$ different mRNA targets; Supplementary Table 3), raising the possibility that muscle tissue in our model is unusually susceptible to antisense silencing. We examined the functional integrity of the muscle

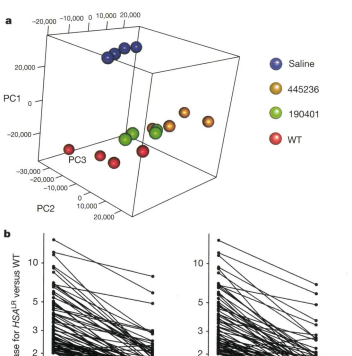


Figure 2 | Effects of ASOs on the transcriptome in quadriceps muscle. **a**, Principle component analysis of microarray data shows segregation of HSA^{LR} (saline) away from wild-type mice in widely separated clusters. ASOs caused HSA^{LR} transgenic mice to cluster nearer to wild-type mice (25 mg kg^{-1} biweekly for 4 weeks; $n = 4$ mice per group). **b**, Of the transcripts upregulated in HSA^{LR} versus wild-type mice (saline), >85% showed complete or partial return to normal expression after treatment with ASOs ($n = 4$ mice per group).

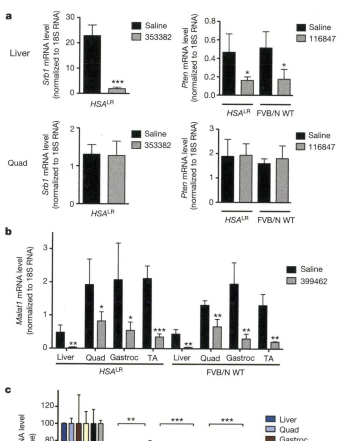


Figure 3 | Differential sensitivity of transcripts to ASO knockdown in skeletal muscle. **a**, In HSA^{LR} or FVB/N wild-type mice, ASOs targeting *Srb1* (353382) or *Pten* (116847) were effective for knockdown in liver but not in quadriceps muscle (qRT–PCR, $n = 4$ per group). Error bars are mean ± s.d. $^*P = 0.02$, $^{***}P < 0.0001$ (*t*-test). **b**, HSA^{LR} and FVB/N wild-type mice were treated with ASO 399462 targeting *Malat1*, a nuclear-retained lncRNA. Levels of *Malat1* transcript in the indicated tissues were determined by qRT–PCR ($n = 4$ ASO, 3 saline). Error bars are mean ± s.d. $^*P = 0.035$, $^{**}P < 0.007$, $^{***}P = 0.001$ for ASO versus saline (*t*-test). **c**, Dose response of *Malat1* knockdown in BALB/c wild-type mice. BALB/c wild-type mice were treated with saline or ASO 399462 targeting *Malat1* at 12.5, 25 and 50 mg kg^{-1} twice per week for 3.5 weeks (7 doses in total; $n = 4$ per group). Tissues were collected for RNA isolation 2 days after the final dose. *Malat1* transcript levels were determined by qRT–PCR. Error bars are mean ± s.e.m. $^*P < 0.01$, $^{**}P < 0.001$, $^{***}P < 0.0001$ (two-way ANOVA).

membrane, a physiological barrier to ASO uptake[24], and found that muscle penetration of the extracellular dye, Evans Blue, was similar in HSA^{LR} and wild-type mice (Supplementary Fig. 11a). Direct analysis of muscle tissue indicated that ASO accumulation was no greater in HSA^{LR} mice than in wild-type controls (Supplementary Fig. 11b, c). Likewise, the mRNA level for RNase H1 was similar in HSA^{LR} and wild-type muscle (Supplementary Fig. 12). We tested ASOs targeting other muscle-expressed transcripts. ASOs for *Pten* phosphatase or *Srb1* (also known as *Scarb1*) scavenger receptor showed efficient target knockdown in liver, but no appreciable knockdown in HSA^{LR} or wild-type muscle (Fig. 3a). Taken together with previous studies, our results indicate specific sensitivity of h*ACTA1*-CUG^{exp} transcripts rather than a general enhancement of ASO activity in HSA^{LR} muscle.

A notable metabolic feature of h*ACTA1*-CUG^{exp} and human *DMPK*-CUG^{exp} mRNA is that processing and polyadenylation are normal but the transcripts are retained in the nucleus[5,6]. Recent studies have shown that RNase H1, the enzyme responsible for antisense knockdown, is localized to the nucleus and mitochondria[14], suggesting that antisense cleavage of nuclear-encoded RNA occurs before nuclear export, and raising the possibility that nuclear-retained transcripts may exhibit enhanced sensitivity. To determine whether other nuclear-retained transcripts show a similar effect we examined metastasis associated lung adenocarcinoma transcript 1 (*Malat1*), an endogenous nuclear lncRNA[9]. We identified MOE gapmer ASOs that produced strong *Malat1* knockdown in cells, in an RNase H1-dependent manner (Supplementary Fig. 13). In wild-type and HSA^{LR} mice, subcutaneous administration of ASOs for 4 weeks caused a greater than 80% reduction in *Malat1* in muscle (Fig. 3b, c), supporting the idea that nuclear-retained transcripts have enhanced sensitivity.

To determine the duration of ASO action in muscle, we examined mice at 15 and 31 weeks after ASO was discontinued, and found that h*ACTA1*-CUG^{exp} knockdown and splicing correction remained strong (not shown). One year after ASO injection was discontinued, target reduction by ASO 190401 had waned, but remained approximately 50% or more for ASO 445236 (Fig. 4a). Even at this late time point the appropriate cleavage products were detected by amplification of complementary DNA 5′ ends (5′ RACE), indicating persistent ASO-RNase H1 activity (Fig. 4b). Consistent with the extent of target reduction, there was partial return of myotonia and splicing defects for

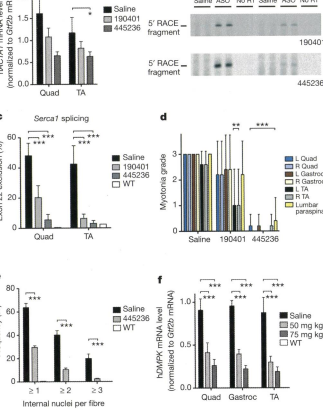


Figure 4 | Duration of ASO activity and *in vivo* targeting of human *DMPK*. **a**–**c**, Two-month-old HSA^{LR} mice received saline or ASO by subcutaneous injection of 25 mg kg^{-1} twice weekly for 4 weeks ($n = 5$ for each ASO, $n = 6$ for saline), with tissues isolated 1 year after the final dose. qRT–PCR analysis of HSA^{LR} transgene mRNA (mean ± s.d.) was normalized to the housekeeping gene *Gtf2b* mRNA (**a**). Results were similar when normalized to total RNA input. 5′ RACE was carried out on muscle RNA obtained 1 week or 1 year after discontinuation of ASO 190401 or 445236 treatment (**b**). PCR products (5′ RACE fragment) migrated at the expected position for ASO•RNase H cleavage products, and were confirmed by DNA sequencing. Quantification of *Serca1* splicing (**c**) and myotonia (**d**) showed a partial return of splicing defects and myotonia for ASO 190401 but not ASO 445236. Myotonia was graded blind by the examiner (**d**). After prolonged knockdown of toxic RNA, the number of internal nuclei per muscle fibre was determined by histologic analysis when mice were aged 14 months (**e**) ($n = 4$ for ASO 445236; $n = 3$ for saline; WT, untreated 3-month-old FVB/N wild-type control mice). **f**, DM328XL mice received subcutaneous injections of saline or ASO 445569 targeting the 3′ UTR of h*DMPK*. The ASO dose was 50 or 75 mg kg^{-1} twice weekly for 4 weeks ($n = 5$, low dose; $n = 4$, high dose; $n = 2$, saline). Tissues were isolated 2 days after the final dose. Dose-dependent reduction of h*DMPK*, normalized to housekeeping gene *Gtf2b* mRNA. Note that h*DMPK* mRNA was undetectable in wild-type mice. No RT, no reverse transcriptase. WT, untreated wild-type littermates of DM328XL transgenic mice ($n = 2$). Error bars are mean ± s.d. *$P < 0.05$, **$P < 0.001$, ***$P < 0.0001$ for ASO-treated versus saline-treated muscle (two-way ANOVA).

ASO 190401, whereas correction by ASO 445236 remained strong (Fig. 4c, d and Supplementary Fig. 14a–e). Furthermore, the persistent knockdown of CUG^{exp} RNA largely prevented the age-dependent myopathic changes in HSA^{LR} muscle, as evidenced by reduced frequency of central nuclei (Fig. 4e) and improved muscle-fibre diameter (mainly a prevention of fibre atrophy) (Supplementary Fig. 15). These findings indicate that ASO activity against h*ACTA1*-CUG^{exp} in muscle is remarkably durable and that long-term reduction of the toxic RNA can protect against structural changes in muscle fibres. Notably, the duration of *Malat1* knockdown in muscle was also prolonged (a greater than 50% reduction at 31 weeks after ASO discontinuation) and more persistent than in liver (Supplementary Fig. 16).

Therapeutic application of this strategy to human DM1 will require transfer of the targeting sequence to h*DMPK*. We developed MOE gapmer ASOs that were active against h*DMPK* in cells. We examined *in vivo* activity after 4 weeks of twice weekly subcutaneous injection in transgenic mice that express h*DMPK* with 800 CUG repeats. The ASO produced significant knockdown of h*DMPK*-CUG^{exp} transcripts in hindlimb muscle (Fig. 4f and Supplementary Figs 17 and 18), supporting the feasibility of silencing the pathogenic *DMPK* allele.

Despite physiological barriers to tissue uptake, our results indicate that systemic targeting of CUG^{exp} RNA is feasible because small amounts of ASOs that enter muscle fibres can hybridize their target and productively engage RNase H1. Although the mechanisms for enhanced sensitivity of CUG^{exp} RNA and *Malat1* are not fully defined, our data suggest that residence time in the nucleus is an important determinant of transcript sensitivity. Features of the nuclear environment that may enhance antisense activity include nuclear localization of RNase H1 (ref. 14) and auxiliary proteins that promote oligonucleotide hybridization[25], and—in the case of CUG^{exp} transcripts—spatial concentration of targets in a small volume[4]. A similar approach may be effective for other genetic disorders that have nuclear accumulation of repeat expansion RNA[26,27]. Previous studies have used CAG-repeat ASOs that bind CUG^{exp} RNA without activating RNase H, in an effort to block the protein interactions or modify the metabolism of the toxic RNA[18,28]. Although this approach was effective with local delivery, initial attempts at systemic delivery were less successful (T.M.W. and C.A.T., unpublished observations), which fits with the expectation that higher tissue concentrations of ASO are required to occupy CUG^{exp} binding sites than to induce RNase H cleavage. Furthermore, the RNase H mechanism is attractive because it exploits the nuclear retention phenomenon to gain a therapeutic advantage, while posing less risk of off-target effects by avoiding a repetitive sequence. Recently, local delivery of RNase H-active CAG-repeat ASOs induced partial CUG^{exp} knockdown, but was accompanied by muscle damage[29], again suggesting that direct targeting of the repeat tract may have pitfalls. Our results also suggest that ASOs are useful for *in vivo* functional characterization and therapeutic modulation of lncRNAs, a large and recently recognized class of regulatory RNAs[30].

METHODS SUMMARY

Experimental mice. All animal experiments were approved by the Institutional Animal Care and Use Committees at the University of Rochester, Genzyme Corporation and Isis Pharmaceuticals.

Subcutaneous injection of ASOs. MOE gapmer ASOs were dissolved in saline and administered by subcutaneous injection in the interscapular region twice per week at the indicated doses.

Quantitative real-time RT–PCR (polymerase chain reaction with reverse transcription) assay. Total RNA was purified from muscle using RNeasy Lipid Tissue Mini Kits (Qiagen). mRNA levels for *ACTA1*, *Srb1*, *Pten*, *Malat1* and RNase H1 were determined on the Applied Biosystems 7500 System using 18S rRNA as normalization control. General transcription factor 2b (*Gtf2b*) and total RNA (Ribogreen assay) served as normalization controls for human *DMPK* and mouse *Dmpk*.

Northern analysis. CUG^{exp} sequences were detected using a ^{32}P end-labelled $(CAG)_8$ DNA oligonucleotide probe.

Electromyography. Electromyography (EMG) was carried out blind under general anaesthesia, as described previously[14].

RT–PCR analysis of alternative splicing. RT–PCR was carried out using the SuperScript III One-Step RT–PCR System with Platinum Taq DNA Polymerase (Invitrogen) and the same gene-specific primers for cDNA synthesis and PCR amplification. PCR products were separated on agarose gels, stained with SybrGreen I Nucleic Acid Gel Stain (Invitrogen) and scanned with a fluorimager.

Transcriptome analysis. Quadriceps-muscle RNA from wild-type or HSA^{LR} transgenic mice treated with vehicle (saline), ASO 445236 or ASO 190401 was processed to cRNA and hybridized on microbeads using MouseRef-8 v2.0 Expression BeadChip Kits (Illumina). Image data were quantified using BeadStudio software (Illumina).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 7 September 2011; accepted 29 June 2012.

1. Harley, H. G. *et al.* Expansion of an unstable DNA region and phenotypic variation in myotonic dystrophy. *Nature* **355,** 545–546 (1992).
2. Buxton, J. *et al.* Detection of an unstable fragment of DNA specific to individuals with myotonic dystrophy. *Nature* **355,** 547–548 (1992).
3. Brook, J. D. *et al.* Molecular basis of myotonic dystrophy: expansion of a trinucleotide (CTG) repeat at the 3′ end of a transcript encoding a protein kinase family member. *Cell* **68,** 799–808 (1992).
4. Taneja, K. L., McCurrach, M., Schalling, M., Housman, D. & Singer, R. H. Foci of trinucleotide repeat transcripts in nuclei of myotonic dystrophy cells and tissues. *J. Cell Biol.* **128,** 995–1002 (1995).
5. Davis, B. M., McCurrach, M. E., Taneja, K. L., Singer, R. H. & Housman, D. E. Expansion of a CUG trinucleotide repeat in the 3′ untranslated region of myotonic dystrophy protein kinase transcripts results in nuclear retention of transcripts. *Proc. Natl Acad. Sci. USA* **94,** 7388–7393 (1997).
6. Mankodi, A. *et al.* Myotonic dystrophy in transgenic mice expressing an expanded CUG repeat. *Science* **289,** 1769–1773 (2000).
7. Bennett, C. F. & Swayze, E. E. RNA targeting therapeutics: molecular mechanisms of antisense oligonucleotides as a therapeutic platform. *Annu. Rev. Pharmacol. Toxicol.* **50,** 259–293 (2010).
8. Geary, R. S. *et al.* Pharmacokinetics of a tumor necrosis factor-alpha phosphorothioate 2′-*O*-(2-methoxyethyl) modified antisense oligonucleotide: comparison across species. *Drug Metab. Dispos.* **31,** 1419–1428 (2003).
9. Wilusz, J. E., Freier, S. M. & Spector, D. L. 3′ end processing of a long nuclear-retained noncoding RNA yields a tRNA-like cytoplasmic RNA. *Cell* **135,** 919–932 (2008).
10. Lorenz, P., Misteli, T., Baker, B. F., Bennett, C. F. & Spector, D. L. Nucleocytoplasmic shuttling: a novel *in vivo* property of antisense phosphorothioate oligodeoxynucleotides. *Nucleic Acids Res.* **28,** 582–592 (2000).
11. Vickers, T. A. *et al.* Efficient reduction of target RNAs by small interfering RNA and RNase H-dependent antisense agents. A comparative analysis. *J. Biol. Chem.* **278,** 7108–7118 (2003).
12. Prasanth, K. V. *et al.* Regulating gene expression through RNA nuclear retention. *Cell* **123,** 249–263 (2005).
13. Wu, H. *et al.* Determination of the role of the human RNase H1 in the pharmacology of DNA-like antisense drugs. *J. Biol. Chem.* **279,** 17181–17189 (2004).
14. Suzuki, Y. *et al.* An upstream open reading frame and the context of the two AUG codons affect the abundance of mitochondrial and nuclear RNase H1. *Mol. Cell. Biol.* **30,** 5123–5134 (2010).
15. Lin, X. *et al.* Failure of MBNL1-dependent post-natal splicing transitions in myotonic dystrophy. *Hum. Mol. Genet.* **15,** 2087–2097 (2006).
16. Osborne, R. J. *et al.* Transcriptional and post-transcriptional impact of toxic RNA in myotonic dystrophy. *Hum. Mol. Genet.* **18,** 1471–1481 (2009).
17. Kanadia, R. N. *et al.* A muscleblind knockout model for myotonic dystrophy. *Science* **302,** 1978–1980 (2003).
18. Wheeler, T. M. *et al.* Reversal of RNA dominance by displacement of protein sequestered on triplet repeat RNA. *Science* **325,** 336–339 (2009).
19. Hasselblatt, P., Hockenjos, B., Thoma, C., Blum, H. E. & Offensperger, W. B. Translation of stable hepadnaviral mRNA cleavage fragments induced by the action of phosphorothioate-modified antisense oligodeoxynucleotides. *Nucleic Acids Res.* **33,** 114–125 (2005).
20. Napierala, M. & Krzyzosiak, W. J. CUG repeats present in myotonin kinase RNA form metastable "slippery" hairpins. *J. Biol. Chem.* **272,** 31079–31085 (1997).
21. Miller, J. W. *et al.* Recruitment of human muscleblind proteins to (CUG)(n) expansions associated with myotonic dystrophy. *EMBO J.* **19,** 4439–4448 (2000).
22. Mankodi, A. *et al.* Expanded CUG repeats trigger aberrant splicing of ClC-1 chloride channel pre-mRNA and hyperexcitability of skeletal muscle in myotonic dystrophy. *Mol. Cell* **10,** 35–44 (2002).
23. Du, H. *et al.* Aberrant alternative splicing and extracellular matrix gene expression in mouse models of myotonic dystrophy. *Nature Struct. Mol. Biol.* **17,** 187–193 (2010).
24. Alter, J. *et al.* Systemic delivery of morpholino oligonucleotide restores dystrophin expression bodywide and improves dystrophic pathology. *Nature Med.* **12,** 175–177 (2006).
25. Pontius, B. W. & Berg, P. Rapid assembly and disassembly of complementary DNA strands through an equilibrium intermediate state mediated by A1 hnRNP protein. *J. Biol. Chem.* **267,** 13815–13818 (1992).
26. Li, L. B. & Bonini, N. M. Roles of trinucleotide-repeat RNA in neurological disease and degeneration. *Trends Neurosci.* **33,** 292–298 (2010).
27. DeJesus-Hernandez, M. *et al.* Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. *Neuron* **72,** 245–256 (2011).
28. Mulders, S. A. *et al.* Triplet-repeat oligonucleotide-mediated reversal of RNA toxicity in myotonic dystrophy. *Proc. Natl Acad. Sci. USA* **106,** 13915–13920 (2009).
29. Lee, J. E., Bennett, C. F. & Cooper, T. A. RNase H-mediated degradation of toxic RNA in myotonic dystrophy type 1. *Proc. Natl Acad. Sci. USA* **109,** 4221–4226 (2012).
30. Wapinski, O. & Chang, H. Y. Long noncoding RNAs and human disease. *Trends Cell Biol.* **21,** 354–361 (2011).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements The work was carried out at the Wellstone Muscular Dystrophy Cooperative Research Center and Center for RNA Biology at the University of Rochester, with support from the US National Institutes of Health (NIH) (grants AR049077, U54NS48843, AR/NS48143, K08NS064293 and U01NS072323), the Saunders Family Neuromuscular Research Fund, Run America and a fellowship (to M.N.) from the Muscular Dystrophy Association and Uehara Memorial Foundation. The authors thank G. Gourdon for providing DM328XL mice, M. Sabripour for assistance with principal components analysis and L. Richardson and S. Leistman for technical assistance.

Author Contributions T.M.W., A.J.L, S.K.P., A.R.M., M.N., S.H.C., B.M.W., C.F.B. and C.A.T. participated in the planning, design and interpretation of experiments. T.M.W., A.J.L., S.K.P., A.R.M and M.N. carried out experiments. T.M.W. and C.A.T. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Readers are welcome to comment on the online version of this article at www.nature.com/nature. The authors declare competing financial interests: details accompany the full-text HTML version of the paper at www.nature.com/nature. Correspondence and requests for materials should be addressed to C.A.T. (charles_thornton@urmc.rochester.edu).

LETTER

doi:10.1038/nature11243

A map of the *cis*-regulatory sequences in the mouse genome

Yin Shen[1]*, Feng Yue[1]*, David F. McCleary[1], Zhen Ye[1], Lee Edsall[1], Samantha Kuan[1], Ulrich Wagner[1], Jesse Dixon[1,2,3], Leonard Lee[1], Victor V. Lobanenkov[4] & Bing Ren[1,5]

The laboratory mouse is the most widely used mammalian model organism in biomedical research. The 2.6×10^9 bases of the mouse genome possess a high degree of conservation with the human genome[1], so a thorough annotation of the mouse genome will be of significant value to understanding the function of the human genome. So far, most of the functional sequences in the mouse genome have yet to be found, and the *cis*-regulatory sequences in particular are still poorly annotated. Comparative genomics has been a powerful tool for the discovery of these sequences[2], but on its own it cannot resolve their temporal and spatial functions. Recently, ChIP-Seq has been developed to identify *cis*-regulatory elements in the genomes of several organisms including humans, *Drosophila melanogaster* and *Caenorhabditis elegans*[3–5]. Here we apply the same experimental approach to a diverse set of 19 tissues and cell types in the mouse to produce a map of nearly 300,000 murine *cis*-regulatory sequences. The annotated sequences add up to 11% of the mouse genome, and include more than 70% of conserved non-coding sequences. We define tissue-specific enhancers and identify potential transcription factors regulating gene expression in each tissue or cell type. Finally, we show that much of the mouse genome is organized into domains of coordinately regulated enhancers and promoters. Our results provide a resource for the annotation of functional elements in the mammalian genome and for the study of mechanisms regulating tissue-specific gene expression.

We identified the genomic localizations of RNA polymerase II (polII), the insulator-binding protein CCCTC-binding factor (CTCF) and three chromatin modification marks, histone H3 lysine 4 trimethylation (H3K4me3), histone H3 lysine 4 monomethylation (H3K4me1) and H3 lysine 27 acetylation (H3K27ac), in 13 adult tissues, four embryonic tissues and two primary cell lines (Fig. 1a, b) by performing chromatin immunoprecipitation followed by high-throughput sequencing (ChIP-Seq)[6] (Supplementary Tables 1 and 2). Enrichment of H3K4me3 or polII binding signals is indicative of an active promoter, whereas the presence of H3K4me1 or H3K27ac outside promoter regions can be used as marks for enhancers[7–11]. CTCF binding is considered a mark for potential insulator elements[12]. In a subset of tissue and cell types, we also performed ChIP-Seq on the co-activator protein p300 and used its promoter-distal binding sites to train an enhancer prediction tool on the basis of chromatin signatures[13]. We determined the transcriptome in each tissue and cell

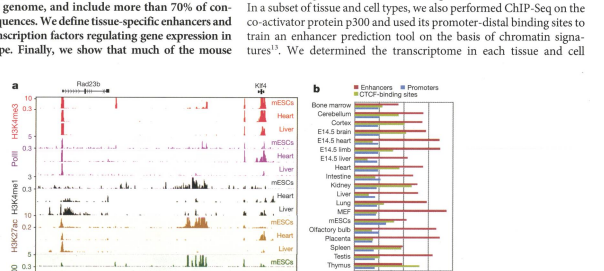


Figure 1 | Identification of *cis*-regulatory elements in the mouse genome. **a**, UCSC genome browser views of ChIP-Seq and RNA-Seq data for mESC, heart and liver (chromosome 4). The values on the *y*-axis for ChIP-Seq data are input normalized intensities. kb, kilobases. **b**, An overview of the predicted regulatory elements in the 19 tissue and cell types. E14.5, embryonic day 14.5; MEF, murine embryonic fibroblast. **c**, Percentages of known *cis*-regulatory elements recovered in this study.

[1]Ludwig Institute for Cancer Research, 9500 Gilman Drive, La Jolla, California 92093-0653, USA. [2]Medical Scientist Training Program, University of California, San Diego School of Medicine, 9500 Gilman Drive, La Jolla, California 92093-0653, USA. [3]Biomedical Sciences Graduate Program, University of California, San Diego School of Medicine, 9500 Gilman Drive, La Jolla, California 92093-0653, USA. [4]Laboratory of Immunogenetics, National Institute of Allergy and Infectious Diseases, Twinbrook I NIAID Facility, Room 1417, 5640 Fishers Lane, Rockville, Maryland 20852, USA. [5]Department of Cellular and Molecular Medicine, Institute of Genomic Medicine, Moores Cancer Center, University of California, San Diego School of Medicine, 9500 Gilman Drive, La Jolla, California 92093-0653, USA.
*These authors contributed equally to this work.

type through RNA-Seq experiments, using a protocol that can detect both the abundance and strand of origin of RNA transcripts[14] (Supplementary Fig. 1). By analysing the genomic occupancy of the above chromatin marks and transcription factors (Supplementary Methods), we identified 295,676 non-redundant *cis*-regulatory sequences, including 53,834 putative promoters, 234,764 potential enhancers and 111,062 CTCF-binding sites (Fig. 1b). With an estimated span of 1,000 base pairs for each element, the combined length of these putative *cis* regulatory sequences is 295.6 million base pairs, or 11% of the mouse genome.

To determine the accuracy and completeness of our *cis*-regulatory sequence mapping, we first compared the identified promoters with known promoters. We recovered 79% of RefSeq-annotated promoters[15] (Fig. 1c and Supplementary Fig. 2a) and confirmed an additional 62% of University of California, Santa Cruz (UCSC)-annotated promoters (13,205 out of 21,433) that are not annotated in RefSeq. As expected, annotated promoters not recovered by our study are generally expressed in tissues that were not investigated in this work (Supplementary Table 3). In addition to the annotated promoters, we also identified 13,438 novel promoters. When tested with a luciferase reporter, 85% of 65 randomly selected novel promoters showed significant promoter activity in at least one orientation ($P < 0.01$, Student's *t*-test) (Supplementary Fig. 3a, b), supporting their function as promoters. Next we compared the predicted enhancers with a list of 726 experimentally validated enhancers[16] and found that 82% of them were correctly identified in this study (Fig. 1c and Supplementary Fig. 2b). We also randomly selected eight predicted murine embryonic fibroblast (MEF) enhancers for validation and found that six of them (75%) gave positive results (Supplementary Fig. 4) ($P < 0.01$, Student's *t*-test), supporting the reliability of our enhancer identification method. In addition, we recovered 94.5% of previously reported CTCF-binding sites in mouse embryonic stem cells (mESCs)[17] (Fig. 1c), demonstrating the high sensitivity of our detection method for CTCF binding. Further, we detected 77,236 novel CTCF-binding sites, 87.5% of which contained the canonical CTCF motifs ($P < 2.2 \times 10^{-16}$, binomial distribution). The novel CTCF-binding sites tend to be more tissue-specific than the sites identified previously (Supplementary Fig. 5). The above evidence indicates that we have correctly identified most known *cis*-regulatory sequences and have uncovered many novel ones.

Functional elements are often under negative selection during evolution, so a high level of sequence conservation is frequently used as evidence of function. However, there are also reports showing that transcription factor binding may be rapidly lost or gained during evolution[18,19], arguing that the usage of *cis*-elements may evolve more quickly. We examined the sequence conservation of different classes of the *cis*-regulatory sequences identified in this study, and found that promoters are characterized by the highest degree of sequence conservation (Fig. 2a). In contrast, CTCF-binding sites and enhancers have a much lower but still significant level of sequence conservation. We next assessed the level of conservation of *cis*-regulatory element usage between the mouse and human genomes in embryonic stem cells (ESCs)[10] (Fig. 2b). More than 70% of homologous promoters are associated with H3K4me3 in both species, confirming a high degree of conservation in promoter usage (Fig. 2c, d). However, only 25.7% and 24.8% of enhancers and CTCF-binding sites, respectively, found in human ESCs are still associated with H3K4me1 or CTCF binding in mESCs, despite a high degree of sequence conservation (Fig. 2c). These results suggest that the *cis*-regulatory elements identified in the mouse genome are under different selective pressure during evolution, with promoters being most conserved in both sequence and usage, whereas enhancers and CTCF-binding sites are undergoing a considerable degree of evolution. This result agrees well with the recent findings of large interspecies differences and divergence of transcriptional regulation[18].

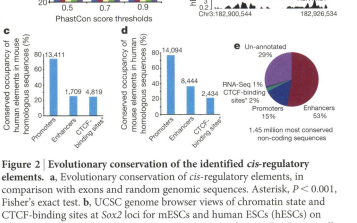


Figure 2 | Evolutionary conservation of the identified *cis*-regulatory elements. **a**, Evolutionary conservation of *cis*-regulatory elements, in comparison with exons and random genomic sequences. Asterisk, $P < 0.001$, Fisher's exact test. **b**, UCSC genome browser views of chromatin state and CTCF-binding sites at *Sox2* loci for mESCs and human ESCs (hESCs) on chromosome 3. DNA sequences, chromatin states and CTCF binding are all conserved in this region. **c**, Number of hESC regulatory elements that are conserved and predicted as regulatory elements in mESCs. **d**, Number of mESC regulatory elements that are conserved and predicted as regulatory elements in hESCs. **e**, Functional annotation of the conserved non-coding sequences based on the *cis*-regulatory elements identified in this study. The asterisk in **c**, **d** and **e** indicates CTCF-binding sites that do not overlap with either promoters or enhancers.

Comparative genomic methods have identified a significant number of mammalian sequences as non-protein coding but undergoing negative selection during evolution, commonly referred to as conserved non-protein-coding sequences (CNSs). These sequences are suspected to have important biological roles, yet their precise function remains to be defined. We compared our map of *cis*-regulatory elements with a list of CNSs[20] and found that 70% of them fall into one of the three classes of predicted *cis*-elements: 15% as promoters, 53% as enhancers and 2% as CTCF-binding sequences. Additionally, 1% of the CNSs seem to be non-coding RNA sequences as supported by the RNA-Seq data (Fig. 2e and Supplementary Fig. 2c). Most CNSs therefore seem to function in regulating transcription.

We previously showed that enhancers in the human genome are associated with active chromatin marks in a cell-type-specific manner, whereas promoter and insulator elements tend to be ubiquitously occupied in multiple cell lines[10]. Here we found that the occupancy of enhancers by H3K4me1 in the mouse genome is still the most tissue-specific (Fig. 3a). In contrast, we observed that whereas H3K4me3 occupies most RefSeq promoters in multiple tissues, a significant number of promoters, especially the novel promoters discovered in this study, show tissue-specific occupancies by H3Kme3 or polII (Fig. 3a) (Supplementary Fig. 3d), with many of them corresponding to alternatively used promoters (Supplementary Table 4 and Supplementary Fig. 6). We also found that most CTCF-binding sites are occupied in multiple tissues (Fig. 3a). The tissue-specific CTCF-binding sites showed significant overlap with enhancers ($P < 1.8 \times 10^{-143}$, binomial distribution), whereas the ubiquitous CTCF-binding sites overlapped significantly with promoters ($P < 9.0 \times 10^{-43}$, binomial distribution) (Supplementary Fig. 5b, c), suggesting that a fraction of the CTCF-binding sites may function through promoters and enhancers, although the exact role of CTCF at these regions remains unclear. These results indicate that a large fraction of *cis*-regulatory elements are active in a tissue-specific manner and are most probably involved in regulating tissue-specific gene expression.

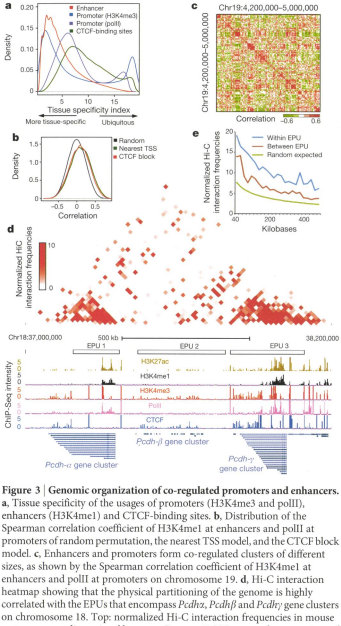


Figure 3 | Genomic organization of co-regulated promoters and enhancers. **a**, Tissue specificity of the usages of promoters (H3K4me3 and polII), enhancers (H3K4me1) and CTCF-binding sites. **b**, Distribution of the Spearman correlation coefficient of H3K4me1 at enhancers and polII at promoters of random permutation, the nearest TSS model, and the CTCF block model. **c**, Enhancers and promoters form co-regulated clusters of different sizes, as shown by the Spearman correlation coefficient of H3K4me1 at enhancers and polII at promoters on chromosome 19. **d**, Hi-C interaction heatmap showing that the physical partitioning of the genome is highly correlated with the EPUs that encompass *Pcdh*α, *Pcdh*β and *Pcdh*γ gene clusters on chromosome 18. Top: normalized Hi-C interaction frequencies in mouse cortex as a two-dimensional heatmap. Bottom: UCSC genome browser views of the same regions, including the identified EPUs and the ChIP-Seq data (H3K27ac, H3K4me1, H3K4me3, polII and CTCF) in cortex. **e**, The average normalized Hi-C interaction frequencies for enhancer–promoter pairs within EPUs, between EPUs, and expected by random chance.

Enhancers are important in regulating tissue-specific expression patterns during mammalian development. However, finding target genes for enhancers is not straightforward because they are frequently distal from the genes they control. Assigning enhancers to the nearest transcription start sites is the most widely used method. A recently published strategy associates enhancers and promoters located within the same domain defined by the CTCF-binding sites, assuming that insulators can block promoter–enhancer interactions[10]. We evaluated these two methods by assessing the Spearman correlation coefficients (SCCs) between H3K4me1 signals at enhancers and the polII intensities at target promoters (Supplementary Methods). As a control, we observed that the SCCs from the randomly paired enhancers and promoters have a bell-shaped distribution with a median of 0 (Fig. 3b). The distribution of the SCCs from enhancer–promoter pairs identified by the nearest transcription start site (TSS) model and CTCF block model are only slightly better than the random control, with medians at 0.11 and 0.08, respectively (Fig. 3b). In addition, 34% and 38% of the enhancer/promoter pairs in the nearest TSS model and the CTCF block model, respectively, are negatively correlated, indicating potentially incorrect promoter assignment. To improve the linking of enhancers to their targets, a logistic regression classifier was recently introduced and shown to perform better than the nearest TSS model[11]. However, this model is still based on the one-to-one relationship between an enhancer and a gene, with a bias towards the nearby genes. It has been reported that a significant fraction of enhancers may not target the nearest promoters[12]. Therefore, to gain a better understanding of enhancer/promoter organization we assessed the correlation of the chromatin state at enhancers and polII occupancy at promoters for each possible pair of elements along a chromosome. We observed that co-regulated promoters and enhancers tend to form clusters with variable sizes (Fig. 3c). We developed an algorithm to detect these local clusters, defined as enhancer–promoter units (EPUs) (Supplementary Methods). Performing this analysis genome-wide, we defined 8,792 EPUs that contained at least one promoter and one enhancer (Supplementary Table 5), encompassing 1,258 million base pairs, or nearly half of the mouse genome. The median enhancer-to-promoter ratio per EPU was 5.67 (Supplementary Table 6), which is consistent with the idea that multiple enhancers may be used to regulate a gene[13]. We confirmed that previously defined enhancer–promoter pairs are frequently located within the same EPU. For example, out of the 2,605 putative enhancer–promoter pairs recently defined in the human genome[21], most of their mouse homologues are found within the same EPU (83.8% observed versus 43% expected; $P < 2.2 \times 10^{-16}$, Fisher's exact test). In addition, each of the four linked enhancer–promoter pairs reported by a recent study[24] was found within the same EPU. Finally, seven locus-control regions for *Hbb* genes were all identified within the same EPU[25].

The discovery of EPUs provides strong evidence that the genome is partitioned into functional domains in which *cis*-regulatory elements are coordinately regulated, whereas elements located in different domains are relatively insulated from each other. This organization is reminiscent of recently identified topological domains, defined by chromatin interactions, in the mammalian genome[26,27]. Indeed, comparison of the EPUs with the higher order chromatin organization shows that physical partitioning of the genome is highly correlated with functional partitioning on the basis of the coordinated activities of *cis*-regulatory sequences (Fig. 3d and Supplementary Fig. 7).

EPUs provide a new approach for associating enhancers with their target genes. Instead of being linked to the nearest genes, an enhancer could be assigned to one or more promoters within an EPU that show significant correlation. To validate the enhancer–promoter relationship predicted by this approach (Supplementary Table 7), we examined long-range looping interactions between the enhancers and promoters, reasoning that true enhancer–promoter target pairs should have higher interaction frequencies than neighbouring non-target sites. We performed chromosome confirmation capture (3C) experiments for five enhancer–promoter pairs predicted to be linked in the cortex but not in mouse ES cells, and two enhancer–promoter pairs predicted not to be linked in either tissue or cell type. The five linked pairs showed enrichment of 3C signals, whereas the two non-linked pairs did not, indicating that the EPU analysis can accurately reveal a enhancer–promoter targeting relationship (Supplementary Fig. 8 and Supplementary Table 8). For a systematic evaluation of the enhancer–promoter pairing relationships as defined by this approach, we examined long-range looping interactions in adult mouse cortex genome-wide by using the Hi-C method[28]. We observed that interactions between predicted enhancer–promoter pairs within the same EPUs occured significantly more frequently than interactions between enhancer–promoter pairs of the same genomic distance but across different EPUs or by random chance (Fig. 3e; $P < 2.2 \times 10^{-16}$, Wilcoxon test). These results suggest that EPUs may help in assigning enhancers to their target promoters.

Mammalian development requires a precise temporal gene expression program that is tightly controlled by transcription factors and *cis*-regulatory elements. The map of *cis*-regulatory sequences now provides a chance for us to analyse the potential mechanisms involved in temporal regulation of gene expression. First, we identified enhancers specific to embryonic and adult brain on the basis of H3K4me1 intensities (Fig. 4a). We observed that the former class was associated with genes expressed in neuron differentiation and neuron development, whereas the latter was associated with genes important for adult brain functions, for example the transmission of nerve impulses (Fig. 4b, c and Supplementary Fig. 9). We made similar observations for stage-specific enhancers in liver and heart (Supplementary Figs 9 and 10).

We also systematically identified potential transcriptional regulators acting on tissue-specific gene expression programs. We first defined 19 groups of tissue-specific enhancers on the basis of H3K4me1 occupancy

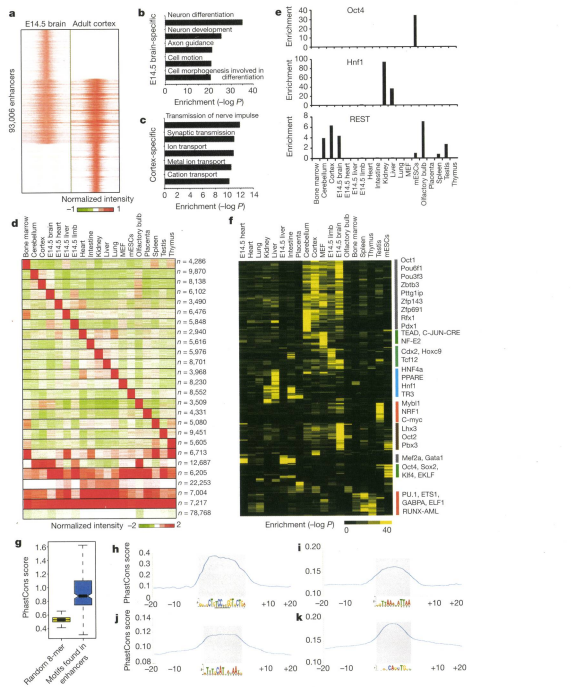


Figure 4 | Motif analysis of tissue-specific enhancers. **a**, Classification of development stage-specific enhancers based on their chromatin state (H3K4me1) between embryonic (embryonic day 14.5; E14.5) and adult brain. **b** and **c**, Gene Ontology analysis for the genes associated with embryonic brain-specific enhancers and adult cortex-specific enhancers. **d**, Classification of tissue-specific enhancers on the basis of their chromatin state (H3K4me1) among different tissue and cell types. The first 19 tissue-specific clusters were used for further motif analysis. The last cluster contains enhancers enriched in multiple tissues with no clear patterns. **e**, Enrichment of three transcription factor recognition motifs in the predicted enhancers. REST, RE1-silencing transcription factor. **f**, Heatmap showing the clustering of 270 transcription factor motifs on the basis of their enrichment in the various groups of enhancers as identified in **e**. **g**, Boxplot showing that the *de novo* motifs found in tissue-specific enhancers are evolutionarily conserved. **h**–**k**, Examples of motifs that show high sequence conservation: **h**, REST motif in cortex-specific enhancers; **i**, Hnf1 motif in kidney-specific enhancers; **j**, Oct4 motif in mESC-specific enhancers; **k**, Atoh1 motif in cerebellum-specific enhancers.

(Fig. 4d). Gene Ontology term analysis confirmed that the enhancers in each group are linked to genes specifically expressed in the corresponding tissue or developmental stage (Supplementary Fig. 11). We also observed that the known motifs of transcription factors that have been reported to function in certain tissues are enriched in the tissue-specific enhancers from the same tissue (Fig. 4e). To identify new transcription factors involved in each group of tissue-specific enhancers, we performed *de novo* motif analysis and identified 206 motifs with a very stringent cutoff ($P < 10^{-20}$; Supplementary Tables 9 and 10). We found that 91% of them (188 out of 206) showed significant levels of evolutionary conservation among the vertebrate species (Fig. 4g, h–k). We annotated the most likely transcription factor for each motif by comparing it with public transcription factor databases and verified that the matching transcription factor was expressed in the corresponding tissue. A total of 62% of the conserved *de novo* motifs (117 out of 188) were associated with a known transcription factor, and 75% of them (88 out of 117) have previously been implicated in the regulation of gene expression in specific tissues (Supplementary Tables 9 and 11). We performed a similar motif analysis for promoters, and compared the top motifs enriched in promoter and enhancer sequences in the same tissue (Supplementary Table 12). Only 11 motifs were shared between the two groups of motifs, whereas 93% of transcription factor motifs enriched in the tissue-specific enhancer were unique only to enhancers, confirming that enhancers and promoters contain different regulatory sequences, as we reported previously[10].

Here we have described an initial survey and a draft annotation of the *cis*-regulatory sequences in the mouse genome. The wide range of tissue and cell types examined in this study provides an unprecedented opportunity to detect tissue-specific and development-specific promoters and enhancers, analyses of which have yielded potential clues to transcription regulators of tissue-specific gene expression programs. We show that nearly half of the mouse genome is organized into EPUs containing enhancers and promoters with correlated activities. These EPUs overlap significantly with recently discovered topological domains, defined by chromatin interactions, thus linking physical partitioning of the genome with transcriptional regulation. Such multigene structures[23,29] probably represent a general feature of genome organization in mammals.

METHODS SUMMARY

Mouse tissues were harvested from eight-week-old male C57Bl/6 mice (Charles River). The murine embryonic fibroblasts were isolated from C57Bl/6 embryos at embryonic day 14.5. ChIP-Seq and RNA-Seq experiments were performed as described[14,30], with the use of Illumina GAIIx and HiSequation (2000) instruments (details are provided in Supplementary Information). Hi-C experiments in adult cortex were conducted as described[24]. A software pipeline to process ChIP-Seq data and predict enhancers is described in Supplementary Methods. Highly correlated biological replicates for ChIP-Seq experiments were pooled for all subsequent data analyses. An algorithm to define the enhancer–promoter unit is given in Supplementary Methods.

Received 12 May 2011; accepted 18 May 2012.
Published online 1 July 2012.

1. Waterston, R. H. *et al.* Initial sequencing and comparative analysis of the mouse genome. *Nature* **420,** 520–562 (2002).
2. Visel, A., Rubin, E. M. & Pennacchio, L. A. Genomic views of distant-acting enhancers. *Nature* **461,** 199–205 (2009).
3. The ENCODE Project Consortium. A user's guide to the encyclopedia of DNA elements (ENCODE). *PLoS Biol.* 9 e1001046 (2011).
4. Gerstein, M. B. *et al.* Integrative analysis of the *Caenorhabditis elegans* genome by the modENCODE project. *Science* **330,** 1775–1787 (2010).
5. Roy, S. *et al.* Identification of functional elements and regulatory circuits by *Drosophila* modENCODE. *Science* **330,** 1787–1797 (2010).
6. Barski, A. *et al.* High-resolution profiling of histone methylations in the human genome. *Cell* **129,** 823–837 (2007).
7. Creyghton, M. P. *et al.* Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl Acad. Sci. USA* **107,** 21931–21936 (2010).
8. Kim, T. H. *et al.* A high-resolution map of active promoters in the human genome. *Nature* **436,** 876–880 (2005).
9. Rada-Iglesias, A. *et al.* A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* **470,** 279–283 (2011).
10. Heintzman, N. D. *et al.* Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* **459,** 108–112 (2009).
11. Heintzman, N. D. *et al.* Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nature Genet.* **39,** 311–318 (2007).
12. Kim, T. H. *et al.* Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. *Cell* **128,** 1231–1245 (2007).
13. Visel, A. *et al.* ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature* **457,** 854–858 (2009).
14. Parkhomchuk, D. *et al.* Transcriptome analysis by strand-specific sequencing of complementary DNA. *Nucleic Acids Res.* **37,** e123 (2009).
15. Pruitt, K. D., Tatusova, T. & Maglott, D. R. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.* **35,** D61–D65 (2007).
16. Visel, A., Minovitsky, S., Dubchak, I. & Pennacchio, L. A. VISTA Enhancer Browser—a database of tissue-specific human enhancers. *Nucleic Acids Res.* **35,** D88–D92 (2007).
17. Chen, X. *et al.* Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* **133,** 1106–1117 (2008).
18. Schmidt, D. *et al.* Five-vertebrate ChIP-seq reveals the evolutionary dynamics of transcription factor binding. *Science* **328,** 1036–1040 (2010).
19. Birney, E. *et al.* Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447,** 799–816 (2007).
20. Siepel, A. *et al.* Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* **15,** 1034–1050 (2005).
21. Ernst, J. *et al.* Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* **473,** 43–49 (2011).
22. Li, G. *et al.* Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. *Cell* **148,** 84–98 (2012).
23. Ong, C. T. & Corces, V. G. Enhancer function: new insights into the regulation of tissue-specific gene expression. *Nature Rev. Genet.* **12,** 283–293 (2011).
24. Kagey, M. H. *et al.* Mediator and cohesin connect gene expression and chromatin architecture. *Nature* **467,** 430–435 (2010).
25. Splinter, E. *et al.* CTCF mediates long-range chromatin looping and local histone modification in the β-globin locus. *Genes Dev.* **20,** 2349–2354 (2006).
26. Nora, E. P. *et al.* Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485,** 381–385 (2012).
27. Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485,** 376–380 (2012).
28. Lieberman-Aiden, E. *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326,** 289–293 (2009).
29. Chepelev, I., Wei, G., Wangsa, D., Tang, Q. & Zhao, K. Characterization of genome-wide enhancer–promoter interactions reveals co-expression of interacting genes and modes of higher order chromatin organization. *Cell Res.* **22,** 490–503 (2012).
30. Hawkins, R. D. *et al.* Distinct epigenomic landscapes of pluripotent and lineage-committed human cells. *Cell Stem Cell* **6,** 479–491 (2010).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank F. Jin, Y. Luu, S. Klugman, A. Y.-J. Kim, Q.-M. Ngo, B. A. Gomez and S. Selvaraj for consultation. The mESC line Bruce4 was a gift from UCSD Transgenic Core. Research funding was provided by the National Human Genome Research Institute (R01HG003991) and the Ludwig Institute for Cancer Research to B.R. Y.S. is supported by a postdoctoral fellowship from the International Rett Syndrome Foundation. J.D. is supported by a pre-doctoral fellowship from the California Institute for Regenerative Medicine.

Author Contributions Y.S., F.Y. and B.R. designed the experiments. Y.S., D.M., Z.Y. and L.L. conducted experiments. F.Y. performed computational analysis. U.W. contributed to RNA-Seq data analysis. J.D. contributed to Hi-C data analysis. S.K. and L.E. performed DNA sequencing and initial data processing. V.L. provided CTCF monoclonal antibodies. Y.S., F.Y. and B.R. prepared the manuscript.

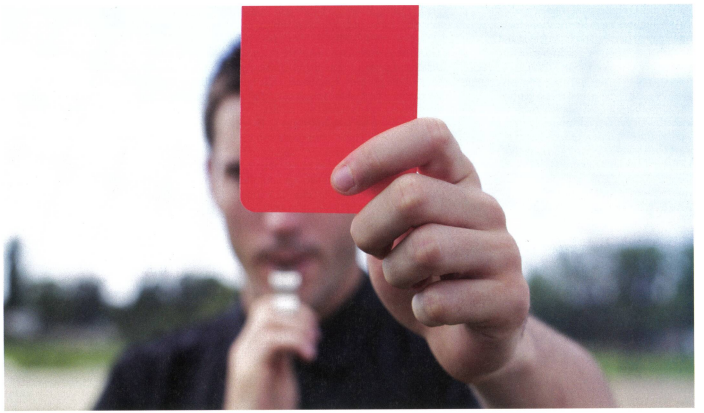
Author Information Data sets are available from the ENCODE website (http://genome.ucsc.edu/ENCODE), the supporting website for this paper (http://chromosome.sdsc.edu/mouse/index.html) and the Gene Expression Omnibus (GSE29184). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to B.R. (biren@ucsd.edu).

CAREERS

GENDER IMBALANCE European universities aim to eliminate inequalities **p.122**

COLLABORATIONS Agreement will send US researchers to Europe **p.122**

NATUREJOBS For the latest career listings and advice **www.naturejobs.com**



STOCK4B CREATIVE/GETTY

COLUMN

The roots of research misconduct

Mentors should understand what causes misconduct among trainees — and keep in mind some possible remedies, argues **William Neaves**.

What is the most effective way for mentors to prevent misconduct among trainees? First, they should make sure that those trainees understand the importance of research integrity. Consistently modelling good practice beats lecturing hands down, and discussing ethical guidelines at laboratory meetings helps the team to appreciate honesty — and the grim consequences of misconduct.

But mentors should also understand the motivation behind some acts of misconduct, and the steps they can take to make sure that misguided trainees don't commit scientific fraud.

While dean of a US biomedical institution more than a decade ago (before my time at the Stowers Institute), I dealt with three cases of scientific misconduct. Each led to an admission of misconduct, sanctions against the perpetrator by the US Office of Research Integrity (ORI) and public disclosure of the person's identity. One case also led to the retraction of several publications. In none of the cases was there any wrongdoing on the part of the mentors.

In the first case, a postdoctoral fellow running 90-minute experiments found that data points tended to plateau after the first 30 minutes. Concluding that nothing of interest happened in the final hour, the postdoc started fabricating those data points. By taking this shortcut, the postdoc quickly generated data, which the mentor incorporated in a manuscript that was then submitted for publication. After belatedly examining the postdoc's lab notebook, the mentor discovered the discrepancy between data collected and data included in figures, and withdrew the manuscript — but not before it had been accepted by a journal. When confronted, the postdoc confessed, and was fired by the host institution. The case took a twist when the postdoc formally accused the mentor of encouraging misconduct by pressuring trainees to generate data. The host institution conducted an inquiry according to ORI standards and found no evidence that other trainees in the lab perceived unusual ▶

GENDER IMBALANCE

Action plans for equality

The 21 members of the League of European Research Universities (LERU) are set to implement strategies to eliminate gender bias in scientific research. *Women, Research and Universities: Excellence Without Gender Bias*, a LERU report published on 10 July, outlines recommendations for universities, funders, policy-makers and publishers to improve gender balance across the European Union. Institutions have agreed to create and launch action plans to mitigate bias, and set up gender-equality offices. A LERU target group will meet this November and annually from next spring to monitor progress, says Kurt Deketelaere, secretary-general of LERU in Leuven, Belgium. The European Commission is using the report in its construction of the European Research Area (see 'Research area first steps').

COLLABORATIONS

US–European deal

US early-career researchers have a new route for in-person collaborations with European colleagues. US National Science Foundation (NSF) postdoctoral fellows or recipients of Faculty Early Career Development awards will be able to spend 6–12 months in teams funded by European Research Council grants under an agreement announced on 13 July. David Stonner, director of the NSF office of international science and engineering in Arlington, Virginia, says that the partnerships could lead to positions for postdocs and help researchers to build and expand networks. "Getting young researchers into European research networks early in their career will pay dividends for years," says Stonner.

EUROPE

Research area first steps

The European Commission on 17 July signed agreements with the first five stakeholders of the European Research Area. The scheme aims to boost research in the European Union (EU) by: coordinating member states, funders and research organizations; making pensions mobile; improving gender equality (see 'Action plans for equality'); and opening hiring practices across borders. EU research is currently fragmented, with little exchange of information and restricted mobility, said Máire Geoghegan-Quinn, European commissioner for research, in a statement.

▶ pressure to produce results.

Fortunately for the scientific reputation of the mentor, he required lab members to maintain bound notebooks that included details of all experiments and data. The case taught him always to scrutinize the relevant notebook entries before submitting a manuscript.

In the second misconduct case, a mentor had asked a postdoc to purify a protein sample so that only a single band remained in a western-blot assay. Instead, the postdoc used a physical mask so that only one band was recorded. A technician found the discarded mask and took it to the mentor, who confronted the postdoc; he admitted falsifying the results. The postdoc was fired by the host institution and sanctioned by the ORI.

What if the technician had not discovered the mask? The mentor could still have taken steps to safeguard the integrity of the work and his own reputation. Having urged the postdoc to purify the sample until a blot showed only one band, he could have sought evidence of the purification steps in the postdoc's lab notebook. And he could have asked a second member of the research team to verify that the results were reproducible.

The final case wreaked havoc on a mentor's research programme. It started when a graduate student falsified a cell-killing assay and fabricated data to support the mentor's favoured hypothesis. The fraud continued when the mentor retained the culprit as a postdoc, on the basis that no one else could "make the assay work properly". Over the course of several years, the postdoc manufactured data for multiple publications.

After the postdoc left the lab to join a biotechnology company, the mentor assigned a new postdoc to perform the assay. When he could not get the expected results, the new postdoc personally paid the former postdoc to perform the assay for him. Sure enough, the results supported the mentor's theory. But the former postdoc would not show the new postdoc how he performed the assay.

The former postdoc then left his biotechnology job, and the mentor rehired him, assigning him responsibility for the assay. But the rehired postdoc would only perform the assay late at night, after everyone had left. Frustrated, the new postdoc hid in the lab one night and saw the culprit pipette a radioactive label directly into scintillation vials, without any attempt to recover it from experimental samples of labelled cells that had been exposed to the (hypothetical) killing agent. The new postdoc reported his observations to the mentor, who immediately informed the dean.

The dean learned that the mentor had never given blinded specimens to the culprit. To avoid having to rely entirely on the new postdoc's testimony, the university's chief academic officer advised the mentor to set up a sting operation. The mentor prepared specimens labelled as experimental that contained no radioactivity. When assayed by the rehired postdoc, these specimens yielded radioactivity that only he could have added to the scintillation vials. When first confronted, he denied everything. The ORI reviewed results of the investigation and concluded that the rehired postdoc had engaged in misconduct. Only then did he acknowledge his guilt. The mentor and co-authors from multiple institutions retracted four high-profile publications that had been based on the fabricated data.

On reflection, what could a mentor have done to prevent this debacle? Simply keeping experimental and control specimens blinded during analysis would have sufficed. Moreover, I believe that a mentor in such circumstances should hear alarm bells if only one person in a lab can get the assay to work. Whenever results depend on human manipulation or measurement, team members should verify each other's work. When a new person joins the lab, the mentor can make it clear that verification practices do not reflect mistrust. For consistency, co-workers should also repeat the mentor's measurements.

"Mentors should not avoid a discussion on research integrity just because of their own discomfort."

Do such practices prevent fraud? They certainly make it more difficult. Just as importantly, they protect against inadvertent error and subconscious bias. Many of us wish for data that support our theories, and trainees may anticipate outcomes that would please the mentor. In general, evidence would suggest that very few trainees curry favour by fabricating data, but mentors should be careful not to encourage misconduct by signalling their disappointment when a trainee's data confound expectations. The chances of falsification or fabrication of results are greatly reduced when a lab uses only blinded specimens and when other lab members are always responsible for independently verifying reproducibility.

In my experience, mentors often avoid discussing scientific misconduct with lab members, perhaps out of a misguided concern that doing so might imply mistrust. There are ways, however, to circumvent this. For example, mentors could broach the topic by first discussing the increasing incidence of retractions (up tenfold in the past decade; see *Nature* **478,** 26–28; 2011). In that way, they can engage trainees without calling into question anyone's integrity.

Mentors should not avoid a discussion on research integrity just because of their own discomfort. The potential consequences for careers and reputations are too severe. ■

William Neaves *is the president emeritus of the Stowers Institute for Medical Research in Kansas City, Missouri.*

European Head Office, London — naturejobs@nature.com
The Macmillan Building, 4 Crinan Street, London N1 9XW, UK
Tel: +44 (0) 20 7843 4961 Fax: +44 (0) 20 7843 4996

Global Head:
Nils Moeller, +44 (0) 20 7843 4953 — n.moeller@nature.com

Southeast UK/RoW:
Emma Hare, +44 (0) 20 7843 4944 — e.hare@nature.com

Northeast UK/Ireland:
Matthew Ward, +44 (0) 20 7014 4059 — m.ward@nature.com

France/Switzerland/Belgium:
Julie Aramburu, +44 (0) 20 7843 4994 — j.aramburu@nature.com

Scandinavia/Spain/Portugal/Italy:
Megan Ball, +44 (0) 20 7843 4973 — m.ball@nature.com

North Germany/The Netherlands/Eastern Europe:
Kristin Doelling, +44 (0) 20 7843 4970 — k.doelling@nature.com

South Germany/Austria:
Maria Kubalova, +44 (0) 20 7014 4084 — m.kubalova@nature.com

Natureevents:
Matthew Clare, +44 (0) 20 7014 4015 — m.clare@nature.com

Business Development Manager, Middle East:
Jon Giuliani, +44 (0) 20 7843 4922 — j.giuliani@nature.com

USA Head Office, New York — naturejobs@us.nature.com
75 Varick Street, 9th Floor, New York, NY 10013-1917
Tel: +1 800 989 7718 Fax: +1 800 989 7103

US Sales Manager:
Ken Finnegan, +1 212 726 9248 — k.finnegan@us.nature.com

Japan Head Office, Tokyo
Chiyoda Building, 2-37 Ichigayatamachi, Shinjuku-ku, Tokyo 162-0843
Tel: +81 3 3267 8751 Fax: +81 3 3267 8746

Asia-Pacific Sales & Business Development Manager:
Yuki Fujiwara, +81-3-3267-8765 — y.fujiwara@natureasia.com

Business Development Manager, Greater China/Singapore:
Gloria To, +852 2811 7191 — g.to@natureasia.com

India
3a, 4th Floor, DLF Corporate Park, Gurgaon, 122 002, India

Advertisement Sales:
Vikas Chawla, +91 124 2881057 — v.chawla@nature.com

NANJING UNIVERSITY

"Deng-feng Scholar" Position: Nanjing University, China

Nanjing University, a key comprehensive university in China under the direct supervision of the Ministry of Education, dates from 1902 when it was known as Sanjiang Normal School. The university invites applications for "Deng-feng Scholar" position in the field of sciences, technology, medicine, social sciences, art and humanity. We particularly welcome creative and energetic individuals with strong backgrounds in all of those fields.

This position offers outstanding scientific resources, reasonable start-up funding, attractive yearly salary and living conditions. The position also offers the opportunity for young scholars to undertake research and studies at the international famous university.

For consideration please download the application form from **http://rczp.nju.edu.cn/urp-publish/publish/XmlFile/1/W_88_313.html** and submit filled form and the required attachments to relative school or department shown in the webpage. The application period is not limited. Questions regarding this position could also be directed to **rcb@nju.edu.cn,** tel. 86-25-83686783

JP227076R

FACULTY APPOINTMENTS

for cardiovascular biomedical researchers

DALHOUSIE UNIVERSITY
Inspiring Minds
Faculty of Medicine

Dalhousie University invites applications to fill two tenure-stream positions at the Assistant or Associate Professor level on the new Faculty of Medicine campus in Saint John, New Brunswick. Successful applicants will be expected to develop an independent laboratory research program in cardiovascular disease including on-site graduate training. You will join two recently arrived researchers and their teams to create a group of at least four researchers to work within a new 5800 sq ft laboratory. It has an open-concept, shared bench and floor space, with separate spaces for microscopy, bacterial and tissue culture, and animal care. You will have the unique and exciting opportunity to help develop and mould this new basic science research program.

Applicants from all basic science disciplines with a cardiovascular focus are invited to apply, including (but not limited to) biochemistry, molecular/cell biology, genetics, anatomy/developmental biology, pharmacology, immunology and physiology. Cardiovascular diseases of interest could include (but are not limited to) atherosclerosis, ischemia, acute coronary syndromes, thrombosis, valvular heart disease, cardiomyopathy, heart failure, arrhythmia and hypertension. Applicants must have a Ph.D. or equivalent and significant postdoctoral experience and have published in peer reviewed journals. Successful applicants will be chosen based on their individual attributes including demonstrated potential to conduct independent biomedical research and to teach at the undergraduate and graduate levels. The selection process will also focus on how the candidate's strengths and interests will fit into a cohesive research group, with strong links to the clinical cardiovascular team at the NB Heart Centre based in Saint John and to researchers at both the University of New Brunswick (UNB) and the Halifax campus of Dalhousie University Faculty of Medicine.

The Dalhousie Medicine New Brunswick program (DMNB; **http://newbrunswick.medicine.dal.ca**) now in its' third year of operation, opened in September 2010, enrolling 30 medical students per year. Successful applicants will make important contributions to the delivery of the Dalhousie medical curriculum. A five-minute walk takes you from the Program's main research and education building on the Saint John campus of UNB, past the health sciences training facility of the New Brunswick Community College, and finally to the Saint John Regional Hospital, a large tertiary care teaching hospital which is home to the NB Heart Centre **(http://www.ahsc.health.nb.ca/Programs/NBHC/).** Saint John is a port city located on the magnificent Bay of Fundy, with a greater metropolitan population of 120,000 **(http://www.tourismsaintjohn.com).**

Successful applicants will be appointed to the appropriate department within the Faculty of Medicine at Dalhousie University, Halifax campus, Nova Scotia **(http://www.medicine.dal.ca)** that will provide support to you in Saint John. You will also have an appointment at the University of New Brunswick **(http://www.unb.ca),** with opportunities to engage in undergraduate and graduate programs on campus.

Applications must be in by November 6th, 2012 to guarantee consideration or until the position is filled. Please send a curriculum vitae, a two-page description of your proposed research program, a brief statement of teaching philosophy and arrange to have 3 letters of reference (2 of which must be academic) sent under separate cover directly to Dr. Frank McCarthy **(g.frankmccarthy@dal.ca)** Anatomist DMNB, Assistant Professor, Dept. of Anatomy and Neurobiology Dalhousie University New Brunswick, c/o Ms. Kelly Sherman, PO Box 5050, 100 Tucker Park Road, Saint John, NB, Canada, E2L 4L5.

All qualified candidates are encouraged to apply; however, Canadians and permanent residents will be given priority. Dalhousie University is an Employment Equity/Affirmative Action employer. The University encourages applications from qualified Aboriginal people, persons with a disability, racially visible persons and women.

NW226833R

ETH

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Assistant Professor of Cellular Biochemistry
Assistant Professor of Systems Biology

The Department of Biology (www.biol.ethz.ch) at ETH Zurich invites applications for above-mentioned Assistant Professorships (non-tenure track).

The Department of Biology offers outstanding scientific opportunities to participate in interdisciplinary research projects, including close interactions with the Swiss Initiative in Systems Biology (SystemsX.ch), the Functional Genomics Center Zurich, and the light- and electron microscopy facilities at ETH Zurich. An interactive, state-of-the-art research environment is provided at the campus Hoenggerberg, with colleagues in computer science, chemistry and physics both at ETH Zurich and University of Zurich.

Research programmes in **Cellular Biochemistry** should be directed towards molecular mechanisms underlying the dynamic organisation of cellular processes and cell architecture with a preferred emphasis on membrane dynamics and lipid biology. Potential topics include, but are not limited to cell growth and division, intracellular transport, and organelle organisation and inheritance. The professorship will ideally be located in the Institute of Biochemistry, taking advantage of its strengths in light microscopy.

Research in **Systems Biology** should focus on cellular systems and processes as integrated networks of interacting molecules. The research programme should be directed to the dynamics of uni- or multicellular systems by applying or developing quantitative methods that may include omics techniques, live imaging or single cell analyses. The professorship will be part of the Institute of Molecular Systems Biology.

The successful candidates will be expected to establish and lead an independent and internationally competitive research group that attracts external funding, with the support of the joint mentoring programme between the Department and the hosting Institutes. In addition, they will actively contribute to the research environment in the Department, and participate in the teaching of graduate and undergraduate students. They will contribute to the ongoing efforts to increase teaching of quantitative approaches in biology. In particular, the candidates will participate in the Bachelor and Master curriculum in Systems Biology and Cellular Biochemistry. The new professors will be expected to teach undergraduate level courses (German or English) and graduate level courses (English).

The assistant professorships have been established to promote the careers of younger scientists. The initial appointment is for four years with the possibility of renewal for an additional two-year period.

Your application should include your curriculum vitae, a list of publications, a statement of your research and teaching interests, and the names of at least three referees. The letter of application should be addressed **to the President of ETH Zurich, Prof. Dr. Ralph Eichler. The closing date for applications is 30 September 2012.** ETH Zurich is an equal opportunity and affirmative action employer. In order to increase the number of women in leading academic positions, we specifically encourage women to apply. ETH Zurich is further responsive to the needs of dual career couples and qualifies as a family friendly employer. **Please apply online at www.facultyaffairs.ethz.ch.**

W227110R

Imperial College London

Junior Research Fellowships

Imperial College London is delighted to invite applications for its Junior Research Fellowship scheme, to give outstanding early-career researchers an opportunity to develop independent research programmes. In 2013, we will be awarding up to 20 Fellowships. Each one offers up to three years funding (stipend in the range £31,300 to £37,800 and up to £10,000 per annum research expenses). Fellowships are designed to provide freedom to focus on research: there are no obligatory teaching or administrative duties. Applicants will be encouraged to apply for additional sources of funding to allow them to strengthen their research programme.

Proposals will be accepted in all research areas active at Imperial College London. Multidisciplinary projects are welcomed.

Applicants will need to identify an Imperial College London academic staff member to act as their sponsor who will provide laboratory space, facilities and mentorship for the tenure of the Fellowship. Candidates may be from the UK or abroad and will be normally expected to have no more than 4 years post-doctoral research experience in an academic environment at the time of application. Proposals will be judged primarily on the basis of scientific merit, track record of the applicant and potential for development of an independent research programme.

Application forms can be found at www.imperial.ac.uk/jrf

Informal enquiries to
juniorresearchfellowships@imperial.ac.uk

Deadline for proposals is October 31 2012 (midnight GMT).

Athena SWAN Bronze Award

Stonewall DIVERSITY CHAMPION

Committed to equality and valuing diversity. We are also an Athena Bronze SWAN Award winner, a Stonewall Diversity Champion and a Two Ticks Employer.

U227171R

Medical Faculty Heidelberg

The **Parasitology Unit** at the **Medical Faculty** of Heidelberg University invites applications for the position of

Full Professor (W3) for Integrative Parasitology

The successful candidate should be a researcher with an excellent international scientific profile and reputation, who is able to represent the subject of Integrative Parasitology in research and teaching. Applicants should be able to participate in the medical and biological teaching activities of the Parasitology Unit with commitment and enthusiasm.

We encourage a thematic focus of the research in the area of „interdisciplinary investigation of parasites" using methods such as molecular biology, cell biology, biophysics and modern imaging analysis. The applicant should hold a PhD in life sciences or medicine. Further requirements are either the "Habilitation", or equivalent scientific and educational experience, and strong records in securing soft money and interdisciplinary collaborative research. For further employment prerequisites please refer to paragraph 47 and 48, section 3, line 4 of the State Law on Higher Education of Baden-Württemberg.

Heidelberg University is committed to strengthen the proportion of women in research and teaching and strongly encourages applications of qualified female scientists. Handicapped applicants with equal qualifications will be given preferential consideration.

Please send your application until **September 3rd 2012** to **Prof. Dr. C. R. Bartram, Dean of the Medical Faculty of Heidelberg University, Im Neuenheimer Feld 672, D-69120 Heidelberg.** The application documents should meet the criteria that can be requested from dekanat@med.uni-heidelberg.de.

W226995R

Westfalian Wilhelms-University Muenster, Germany

The Medical Faculty of the Westfälische Wilhelms-Universität Münster offers the position of a

Full Professor (W3) of 'Translational Oncology'

The Medical Faculty of the Westfälische Wilhelms-Universität (WWU) Münster invites applications from medical doctors for a position in the newly established 'Cells in Motion - CiM' excellence cluster available immediately. The professorship will be a 5-year appointment initially with tenure-track option. The position will be supplemented by scientific, clinical and technical staff positions and equipment.

This professorship is one of the clinical translational professorships promoted by CiM, a new interfaculty excellence cluster involving Medicine, Physics, Mathematics and Computer Sciences, Chemistry and Pharmacy and Biology (www.cells-in-motion.de). It aims at bridging existing excellence in Münster in single cell imaging and molecular whole body imaging. Clinical translational professors dedicate 60 % or more of their time to translational research activities in addition to having reduced clinical duties. The professor of 'Translational Oncology' will head an independent patient care unit in the Dept. of Haemotology and Oncology (Director: W. Berdel) at the university hospital in addition to her/his independent research lab. The research focus of the applicant should synergize with one or more of the special interests of the department in a) genetic and epigenetic control of tumorigenesis, b) vascular targeting, or c) cellular immunotherapy. Of particular interest are scientists with expertise in analysing dynamics of tumour cells in vivo by novel molecular and cellular imaging tools, in a translational setup from models to patients. The successful candidate (male/female) must have an internationally recognized research career and a commitment to innovative teaching.

Preference will be given to individuals whose interests are synergistic with ongoing collaborative research programs at the involved faculties, the Max Planck Institute for Molecular Biomedicine, and the European Institute for Molecular Imaging (EIMI) - such as SFB 629 Molecular Cell Dynamics; SFB 656 Molecular Cardiovascular Imaging; SFB 1009 Breaking Barriers; SFB/TR 128 Multiple Sclerosis, existing graduate schools, and the Interdisciplinary Centre for Clinical Research (IZKF).

Prerequisite for the application are scientific achievements as a junior or full professor, from a postdoctoral lecture qualification (habilitation), as a research scientist at a school of higher education/university, at non-university institutes, industry, administration, or other fields of society within or outside Germany.
As an equal opportunity employer, the WWU is committed to the advancement of individuals without regard to ethnicity, religion, sex, age, disability or any other protected status. Applications of women are specifically invited. In the case of similar qualification, competence, and specific achievements, women will be considered on preferential terms within the framework of the legal possibilities. Handicapped candidates with equivalent qualifications will be given preference.

Electronic submission of CV, documentation of scientific career, publications, acquired third-party funds, the six most important publications and a teaching concept is encouraged. Documents should be submitted to: **Prof. Dr. Wilhelm Schmitz, dekanmed@uni-muenster.de, Dean of the Medizinische Fakultät der Westfälischen Wilhelms-Universität Münster, Albert-Schweitzer-Campus 1, Gebäude D3, 48149 Münster, Germany,** by **30.09.2012.**

W227180R

naturejobs

Need help with your career?
Ask the Expert on naturejobs.com

- **Browse past Q&A**
- **Get professional advice**
- **Email your questions**

Ask the Expert is located in the **Career Toolkit** section of **naturejobs.com**

TUM Technische Universität München

The **TUM Department of Physics** invites applications for a

Full Professorship »Theoretical Physics: Theory of Complex Biosystems«

Pay-scale grade W3, to be filled as soon as possible.

We are looking for a scholar of distinction with a high international reputation in Theoretical Physics: Theory of Complex Biosystems. The successful candidate will have demonstrated an internationally recognized research program as well as outstanding academic records and is expected to perform cutting-edge research in the area of theoretical biophysics. Relevant areas of expertise may include all aspects which address the understanding, design and manipulation of functional biosystems at the molecular or cellular level. The research topics should be related to the structure and dynamics of biosystems, non-equilibrium processes, phase transitions in biosystems or the development of novel theoretical methods and applications to relevant bio-nanosystems. Close collaboration within the interdisciplinary research centers of the Technische Universität München is expected. The Munich area offers a rich collaborative research environment. Of particular interest for the position to be filled are the collaborative research center "Forces in biomolecular systems" (SFB863) as well as the research clusters of Excellence "Nanosystems Initiative Munich" and the "Center for Integrated Protein Science".

A proven ability to attract competitive national and international funding is expected.

In addition to providing strong, committed leadership and inspiration in research, the candidate will also be expected to actively contribute to the ongoing development of the university's research and teaching program, support the promotion of young scientists and participate in university boards and committees.

Teaching assignments include courses in the subject area and the basic courses offered by the department as well as courses for other academic TUM departments. Prerequisites for this position are a university degree, a doctoral degree, teaching skills at university level, and additional academic achievements. Candidates for this position should be aged 51 or under at the time of appointment. The ability to teach in English is a prerequisite for TUM Professors.

As part of the Excellence Initiative of the German federal and state governments, TUM has been pursuing the strategic goal of substantially increasing the diversity of its faculty. As an equal opportunity and affirmative action employer, TUM explicitly encourages nominations of and applications from women as well as from all others who would bring additional diversity dimensions to the university's research and teaching strategies. Preference will be given to disabled candidates with essentially the same qualifications. The TUM Munich Dual Career Office provides support for dual career couples and families.

Applications accompanied by supporting documentation (CV, certificates, credentials, list of publications including 3 selected reprints, list of courses taught, presentation of research and teaching strategies, third-party funding, names and addresses of at least 3 references) should be submitted by **October 15th, 2012** to:

Dekan, Department of Physics
Technische Universität München
James-Franck-Straße 1
85748 Garching, Germany
E-Mail: dekanat@ph.tum.de

W226487R

"Thank you for assisting in advertising the 6x 'Research Associate' positions on the Naturejobs website. I was advised thoroughly on the inventory that would be the best for our organisation and the web inventory worked. We received a huge response and will contact you for future advertisements."

Nanda Rahman, Bio Nano Consulting, UK

naturejobs

Westfalian Wilhelms-University Muenster, Germany

The Medical Faculty of the Westfälische Wilhelms-Universität Münster offers the position of a

Full Professor (W3) of 'Translational Regenerative Medicine'

The Medical Faculty of the Westfälische Wilhelms-Universität (WWU) Münster invites applications from medical doctors for a position in the newly established 'Cells in Motion - CiM' excellence cluster available immediately. The professorship will be a 5-year appointment initially with tenure-track option. The position will be supplemented by scientific, clinical and technical staff positions and required equipment.

This professorship is one of four clinical translational professorships promoted by CiM, a new interfaculty excellence cluster involving Medicine, Physics, Mathematics and Computer Sciences, Chemistry and Pharmacy and Biology (www.cells-in-motion.de). It aims at bridging existing excellence in Münster in single cell imaging and molecular whole body imaging. Clinical translational professors dedicate 60 % or more of their time to translational research activities in addition to having reduced clinical duties. The professor of 'Translational and Regenerative Medicine' will head an independent patient care unit in the Centre of Musculoskeletal Medicine at the university hospital in addition to her/his independent research lab. The research focus of the applicant should synergize with one or more of the special interests of the department in regenerative osteology, endocrinology, immunological and age-related post traumatic and degenerative processes, and development and application of imaging for assessing bone loss and regeneration. The successful candidate (male/female) must hold a German medical license or foreign equivalent and should be board certified in either orthopedic or trauma surgery or internal medicine or related fields. The candidate must have an internationally recognized research career and a commitment to innovative teaching.

Preference will be given to individuals whose interests are synergistic with ongoing collaborative research programs at the involved faculties, the Max Planck Institute for Molecular Biomedicine, and the European Institute for Molecular Imaging (EIMI) - such as SFB 629 Molecular Cell Dynamics; SFB 656 Molecular Cardiovascular Imaging; SFB 1009 Breaking Barriers; SFB/TR 128 Multiple Sclerosis, existing graduate schools, and the Interdisciplinary Centre for Clinical Research (IZKF).

Prerequisite for the application are scientific achievements as a junior or full professor, from a postdoctoral lecture qualification (habilitation), as a research scientist at a school of higher education/university, at non-university institutes, industry, administration, or other fields of society within or outside Germany. As an equal opportunity employer, the WWU is committed to the advancement of individuals without regard to ethnicity, religion, sex, age, disability or any other protected status. Applications of women are specifically invited. In the case of similar qualification, competence, and specific achievements, women will be considered on preferential terms within the framework of the legal possibilities. Handicapped candidates with equivalent qualifications will be given preference.

Electronic submission of CV, documentation of scientific career, publications, acquired third-party funds, the six most important publications and a teaching concept is encouraged. Documents should be submitted to: **Prof. Dr. Wilhelm Schmitz, dekanmed@uni-muenster.de, Dean of the Medizinische Fakultät der Westfälischen Wilhelms-Universität Münster, Albert-Schweitzer-Campus 1, Gebäude D3, 48149 Münster, Germany,** by **30.09.2012.**

W227177R

MANCHESTER 1824
The University of Manchester

Institute of Inflammation and Repair

Research Associates - 2 posts

£29,249 - £30,122 p.a.

We welcome applications from outstanding candidates with a proven track research record in the area of biomaterials and/or polymer science. The posts are tenable for 12 months in the first instance, starting on 1 September 2012.

Post 1 **Ref: M&HS-01514**

You will focus on the development of techniques for the surface biofunctionalisation of porous polymeric constructs (electrospun fibres) and for the preparation of different forms of cell-containing hydrogels.

You should have an undergraduate degree or equivalent in chemistry or materials science and a PhD or equivalent in the area of biomaterials. You will also have a demonstrated research experience at the interface between materials and polymer science, with a good background in organic chemistry and in the use of a variety of characterization techniques.

Post 2 **Ref: M&HS-01512**

You will focus on the evaluation of cell interactions with biofunctional porous polymeric scaffolds, and on the factors influencing matrix remodelling.

You should have an undergraduate degree or equivalent in a biomedical/life sciences discipline and a PhD or equivalent in the area of biomaterials or cell/matrix interactions/tissue engineering. You will also have a demonstrated research experience at the interface between materials and cell biology, with a good background in molecular biology and in materials characterization techniques.

Informal enquiries may be made to Prof. Nicola Tirelli at nicola.tirelli@manchester.ac.uk or on 0161 275 2480.

For further information and to apply for this vacancy please visit our website www.manchester.ac.uk/jobs

If you are unable to apply online please request an application form by emailing hrrecruitment@manchester.ac.uk or calling 0161 275 8835 quoting the reference number.

Closing date: 16 August 2012.

The University will actively foster a culture of inclusion and diversity and will seek to achieve true equality of opportunity for all members of its community.

www.manchester.ac.uk/jobs

U227172R

京都大学生命科学研究科
高次応答制御学講座シグナル機能学分野教授候補者の公募

1. **任用条件**　生命現象をシステムとして理解することを目指す卓越した研究を展開し、優れた教育指導能力を有する者。なお、当該ポストは薬学部における学部教育を兼担する予定である。また、当該ポストは共同研究室として複数設もしくは数階1名の計2名の教員を採用できる予定である。
2. **応募締切**　平成24年 10 月1日（月）必着
3. **提出書類**（A4判）
   （1）履歴書　（2）著書および学術論文目録　（3）主要著書および学術論文別刷（5編以内）
   上記（2）の目録の該当番号に○印を付してください。
   （4）現在までの研究の概要と生命科学研究科における将来の研究・教育の方向（2,000字程度）
   （5）推薦書または本人について評価できる研究者の氏名（2名以上）と連絡先
   （6）過去5年間の競争的資金の獲得状況を記した書類
4. **任用予定時期**　平成25年のなるべく早い時期
5. **男女共同参画**　京都大学は男女共同参画を推進しています。
   多数の女性研究者の積極的なご応募を期待します。
6. **提出先**　〒606-8501　京都市左京区吉田近衛町　京都大学大学院生命科学研究科総務掛
   TEL 075-753-9221
   （郵送の場合は簡易書留とし「高次応答制御学講座教授応募書類在中」と朱書き願います）
7. **問合せ先**　京都大学大学院生命科学研究科高次生命科学専攻
   高次応答制御学講座教授候補者選考委員会　委員長　根岸 学
   TEL 075-753-4547

JP227121RL

“Based on the huge response to our adverts in *Naturejobs* we have hired many excellent candidates annually. We highly appreciate the kind and professional support of the *Nature* team.”

Karin Osibow,
Program Administrator,
Medical University of Graz, Austria

naturejobs

Visit
www.naturejobs.com
to seriously improve
your career prospects.

naturejobs
making science work

Director, National Institute of General Medical Sciences, National Institutes of Health

The NIH is the center of medical and behavioral research for the Nation – – – making essential medical discoveries that improve health and save lives.

Are you a top-level Scientific Researcher or Scientific Administrator seeking a career at one of the preeminent health institutes in the nation and the world? Are you at that point in your career where you're ready to "give back"? The position of Director, National Institute of General Medical Sciences (NIGMS), National Institutes of Health (NIH), offers a unique and exciting opportunity to provide leadership, administer, foster, and support research in the basic and general medical sciences, and in related natural or behavioral sciences. The Director, NIGMS, develops Institute goals, priorities, policies, and program activities, and keeps the Director, NIH abreast of NIGMS developments, accomplishments, and needs as they relate to the overall mission of the NIH. S/he is responsible for managing a high-level complex organization and serves as the chief visionary for the Institute. The mission of NIGMS is to support research that increases understanding of life processes and lays the foundation for advances in disease diagnosis, treatment, and prevention.

We are looking for applicants with senior level experience who have a commitment to excellence and the energy, enthusiasm, and innovative thinking necessary to lead a dynamic and diverse organization.

The successful candidate for this position will be appointed at a salary commensurate with his/her qualifications. Full Federal benefits will be provided including leave, health and life insurance, long-term care insurance, retirement, and savings plan (401k equivalent).

If you are ready for an exciting leadership opportunity, please see the detailed vacancy announcement at http://www.jobs.nih.gov (under Executive Jobs).
Applications will be reviewed starting October 1, 2012 and will be accepted until the position is filled.

NIGMS, NIH AND DHHS ARE EQUAL OPPORTUNITY EMPLOYERS

NW226832R

College of
Biological Sciences

UNIVERSITY OF MINNESOTA
Driven to Discover℠

The College of Biological Sciences at the University of Minnesota is hiring 16 scientists to form six interdisciplinary clusters in emerging areas of biology that connect with other STEM disciplines. All positions are tenure track. Successful candidates will have expertise that complements current faculty and be committed to graduate and undergraduate education.

Research cluster themes include **cellular biophysics**, **genomic variation**, **functional proteomics**, **microbial systems and synthetic biology**, **fungal evolution** and **theoretical biology**.

Learn more at
z.umn.edu/cbsclusterhiring

The University of Minnesota is an equal opportunity educator and employer.

NW226865R

Westfalian Wilhelms-University Muenster, Germany

The Medical Faculty of the Westfälische Wilhelms-Universität Münster offers the position of a

Full Professor (W3) of 'Cardiovascular Organogenesis and Regeneration'

The Medical Faculty of the Westfälische Wilhelms-Universität (WWU) Münster invites applications for a position in the newly established 'Cells in Motion - CiM' excellence cluster available immediately. The position will be supplemented by scientific and technical staff positions and required equipment.

The professorship will be situated in CiM, a new interfaculty excellence cluster involving Medicine, Physics, Mathematics and Computer Sciences, Chemistry and Pharmacy and Biology (www.cells-in-motion.de). It aims at bridging existing excellence in Münster in single cell imaging and molecular whole body imaging. Of particular interest are scientists with expertise in experimental approaches integrating modern developmental and cell biology with advanced imaging. The successful candidate (male/female) must have an internationally recognized research career and a commitment to innovative teaching.

Preference will be given to individuals whose interests are synergistic with ongoing collaborative research programs at the involved faculties, the Max Planck Institute for Molecular Biomedicine, and the European Institute for Molecular Imaging (EIMI) - such as SFB 629 Molecular Cell Dynamics; SFB 656 Molecular Cardiovascular Imaging; SFB 1009 Breaking; SFB/TR 128 Multiple Sclerosis, existing graduate schools, and the Interdisciplinary Centre for Clinical Research (IZKF).

Prerequisite for the application are scientific achievements as a junior or full professor, from a postdoctoral lecture qualification (habilitation), as a research scientist at a school of higher education/university, at non-university institutes, industry, administration, or other fields of society within or outside Germany.
As an equal opportunity employer, the WWU is committed to the advancement of individuals without regard to ethnicity, religion, sex, age, disability or any other protected status. Applications of women are specifically invited. In the case of similar qualification, competence, and specific achievements, women will be considered on preferential terms within the framework of the legal possibilities. Handicapped candidates with equivalent qualifications will be given preference.

Electronic submission of CV, documentation of scientific career, publications, acquired third-party funds, the six most important publications and a teaching concept is encouraged. Documents should be submitted to: **Prof. Dr. Wilhelm Schmitz, dekanmed@uni-muenster.de, Dean of the Medizinische Fakultät der Westfälischen Wilhelms-Universität Münster, Albert-Schweitzer-Campus 1, Gebäude D3, 48149 Münster, Germany,** by **30.09.2012.**

W227179R

www.hud.ac.uk/jobs

Senior Lecturer in Biochemistry

School of Applied Sciences
£36,271 - £45,908
Huddersfield
Full-time, 1.0 FTE
Ref: 6401

We provide an exciting, stimulating place to learn and work. Our standards of academic achievement are high, our environment is inspiring and we embrace innovation in all that we do. And with exceptional opportunities for staff development, our students are not the only ones who thrive here.

The thriving, successful and expanding Department of Chemical & Biological Sciences is seeking an enthusiastic individual to play a major role in the development and teaching of undergraduate and postgraduate courses in Biological Sciences and to undertake an active research programme. You should have a research degree and preferably relevant post-doctoral or industrial experience with a good record of publications and the ability to attract research funding.

For further details and an application form visit www.hud.ac.uk/jobs
Alternatively contact the Personnel Office on 01484 472845.

Closing date: 10 September 2012.
Interview date: 1 October 2012.

Innovative University.
Inspiring Employer.

University of HUDDERSFIELD

U227176R

The Tom Wahlig Foundation was established in 1998 with the aim of bringing together all those concerned with Hereditary Spastic Paraplegia (HSP) and promoting research into both neuroscience as well as clinical care for HSP patients.

TWS TOM WAHLIG STIFTUNG

The foundation is proud to announce second time the

Tom Wahlig Advanced Scholarship
for research into Hereditary Spastic Paraplegia.
The sum of 100.000 €

will be awarded to a distinguished scientist with an excellent record in neurobiology and/or human genetics. Proposals should represent a comprehensive approach focusing on HSP etiology and opening up new therapeutic perspectives. All proposals will be evaluated by an international jury consisting of Prof. Dr. Hans Schöler (MPI Muenster); Prof. Dr. J. Fischer (Humangenetik Freiburg); Prof. Dr. E.B. Ringelstein (Neurologie UKM, Münster); Prof. Dr. Michael Sendtner (Klinische Neurobiologie, Würzburg).

Applicants should submit:

- a five-page exposé for a three-year research project
- their full CV (resumé)
- a two-page summary of their research record and a list of key publications

Applicants have to be submitted by E-Mail to **wahlig@t-online.de** no later **than October 31, 2012.** Detailed information about the foundation and the last winner can be found at **www.hsp-info.de**

W227005AM

Westfalian Wilhelms-University Muenster, Germany

The Medical Faculty of the Westfälische Wilhelms-Universität Münster offers the position of a

Full Professor (W3) of 'Cell mechanics'

The Medical Faculty of the Westfälische Wilhelms-Universität (WWU) Münster invites applications for a position in the newly established 'Cells in Motion - CiM' excellence cluster available immediately. The position will be supplemented by scientific and technical staff positions and required equipment.

The professorship will be situated in CiM, a new interfaculty excellence cluster involving Medicine, Physics, Mathematics and Computer Sciences, Chemistry and Pharmacy and Biology (www.cells-in-motion.de). It aims at bridging existing excellence in Münster in single cell imaging and molecular whole body imaging. Of particular interest are scientists with expertise in experimental approaches integrating modern developmental and cell biology with advanced imaging. The successful candidate (male/female) must have an internationally recognized research career and a commitment to innovative teaching.

Preference will be given to individuals whose interests are synergistic with ongoing collaborative research programs at the involved faculties, the Max Planck Institute for Molecular Biomedicine, and the European Institute for Molecular Imaging (EIMI) - such as SFB 629 Molecular Cell Dynamics; SFB 656 Molecular Cardiovascular Imaging; SFB 1009 Breaking; SFB/TR 128 Multiple Sclerosis, existing graduate schools, and the Interdisciplinary Centre for Clinical Research (IZKF).

Prerequisite for the application are scientific achievements as a junior or full professor, from a postdoctoral lecture qualification (habilitation), as a research scientist at a school of higher education/university, at non-university institutes, industry, administration, or other fields of society within or outside Germany. As an equal opportunity employer, the WWU is committed to the advancement of individuals without regard to ethnicity, religion, sex, age, disability or any other protected status. Applications of women are specifically invited. In the case of similar qualification, competence, and specific achievements, women will be considered on preferential terms within the framework of the legal possibilities. Handicapped candidates with equivalent qualifications will be given preference.

Electronic submission of CV, documentation of scientific career, publications, acquired third-party funds, the six most important publications and a teaching concept is encouraged. Documents should be submitted to: **Prof. Dr. Wilhelm Schmitz, dekanmed@uni-muenster.de, Dean of the Medizinische Fakultät der Westfälischen Wilhelms-Universität Münster, Albert-Schweitzer-Campus 1, Gebäude D3, 48149 Münster, Germany,** by **30.09.2012.**

W227182R

"When trying to recruit good scientists for such a remote institutions like ours, in the Canary Is., the high visibility and wide-range exposure that Naturejobs provides has been crucial. I'm very happy to have international candidates who applied to our position from places ranging between St. Petersburg and Cincinnati. Thanks a lot!"

Eduardo Salido, MD, PhD,
Hospital Universitario de Canarias

nature jobs

REPROGRAPHICS: CTT, London E17 6BU UK and **The Charlesworth Group,** Wakefield, UK.
PRINTED BY: William Gibbons & Sons Ltd, UK; **Publishers Press**, Lebanon Junction, Ky, USA and **Obun Printing Co. Inc,** Tokyo, Japan

a nature conference

Nature Genetics and the Wellcome Trust present:

The Genomics of Common Diseases 2012

Nature Genetics in association with the Wellcome Trust is pleased to announce the sixth annual Genomics of Common Diseases meeting, which will address the following topics across a range of common diseases: exome and whole genome sequencing, clinical and targeted sequencing, regulatory and functional genomics, cancer genomics, clinical translation and pharmacogenetics, statistical analyses of sequencing datasets, population and evolutionary genetics and infectious disease and host genetics.

September 19-22, 2012
Bolger Center, Potomac, MD, USA

KEYNOTE SPEAKERS

Francis Collins (National Institutes of Health, USA)
Thomas Hudson (Ontario Institute for Cancer Research, Canada)
Neil Risch (University of California, San Francisco, USA)
Sarah Tishkoff (University of Pennsylvania School of Medicine, USA)

SPEAKERS

Antonis Antoniou
John Blangero
Mary Carrington
Stephen Chanock
Paul de Bakker
Manolis Kellis
Matthew Meyerson
Patricia Munroe
Benjamin Neale
Christopher O'Donnell
Alkes Price
Jonathan Pritchard
Dan Roden
Nicholas Schork
Christine Seidman
Jay Shendure
Adam Siepel
Pamela Sklar
Nicole Soranzo
Shamil Sunyaev
Peter Visscher
George Weinstock
Cisca Wijmenga
Ramnik Xavier

ORGANIZERS:

Orli Bahcall (*Nature Genetics*, USA)
Aravinda Chakravarti (Johns Hopkins University School of Medicine, USA)
Peter Donnelly (University of Oxford, UK)
Michael Dunn (Wellcome Trust, UK)
Michel Georges (University of Liège, Belgium)
Teri Manolio (National Human Genome Research Institute, National Institutes of Health, USA)
Mark McCarthy (University of Oxford, UK)
Nazneen Rahman (Institute of Cancer Research, UK)
Mary Relling (St. Jude Children's Research Hospital, USA)

For more information and to register visit:
www.nature.com/natureconferences/gcd2012

nature genetics

wellcome trust

DEAD MEAT

Interactive art.

BY POLENTH BLAKE

True art is seamless. That's why I crafted my new gallery out of meat. Patrons squelched around the entrance hall in plastic coats and boots. One man touched the walls gingerly. His hand came away red with blood.

"Welcome to *Dead Meat II*," a voice greeted him from the wall. The voice holes lined the entrance room, each with its own set of vocal cords and lungs. They'd been modelled after mine. Although the tone was right, the organic quality was missing. Higher functions needed the computer's control and it never quite got it.

The entrance had information on the original *Dead Meat*. It had been a stroke of genius, if I do say so myself. Meat bricks around a metal frame. I'd fused the meat together with transglutaminase, so it looked like it'd grown as one solid lump.

Genius, but temporary. All meat goes bad in the end. Still, Clarissa Brookes was a worldwide name before the last piece rotted away. I had no trouble getting sponsors for this gallery.

The door to the first section opened as I approached. Far superior to any mechanized door. The muscles could open it slowly for the casual wanderer or pull back quickly when a child ran through. The door understood. Cold metal and plastic couldn't mimic that.

My sponsors had insisted on displays of my older work. I'd resisted at first, but I came around to seeing the merit. My patrons learnt with me. Paintings of daisies growing out of women. An elephant's head flowed into a whale's body. The works were crude, but the concept of seamless design was lurking in the brush strokes.

"They're a bit creepy, aren't they?" a woman noted to her companion.

I resisted commenting. Some patrons were slower learners than others.

For my first sculpture, I'd sewn the front ends of two sheep together. The fleece was arranged to cover the seams, but I knew that they were there. Stitches are still stitches, no matter how well you hide them.

Blood oozed from the ceiling more freely in the next room. A few splashes landed on the hood of one patron. After a whispered conversation with his partner, they headed back towards the entrance.

➲ **NATURE.COM**
Follow Futures on Facebook at:
go.nature.com/mtoodm

I pitied them for their squeamishness. I'd refused the plastic coverings, preferring to wander the halls in a mini skirt and tank top. I wanted to feel the blood on my skin.

Most of my middling work was inconsequential. The same old designs, as I struggled to find my way. I ignored the pieces in favour of the patrons. A child watched the floor, pressing his foot against the flesh to see it ooze. He realized that the true art was the building. It made it all worthwhile.

A man fainted and nicked the wall with his watch as he slumped into it. Pores opened in the floor to reabsorb the blood from the cut, while the edges of the tear pulled together. Designing the repair nanobots had been the hardest part. A true immune system would react with or without control, but the nanobots needed the computer. They were the final seam in the construction. For the next *Dead Meat*, I'd do away with the last of the machinery.

The doors jerked open. Two security guards hurried through, in the manner of people who don't want to panic the patrons. The door muscles trembled in response.

I followed the guards, resting a reassuring hand on the doorframe as I passed.

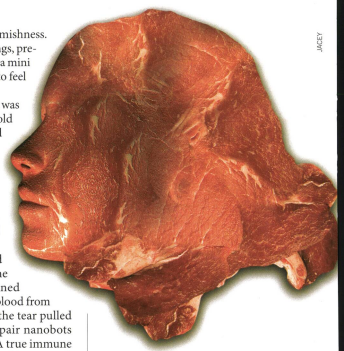
Muffled shouts came from farther in the gallery. No doubt one of the protestors, spouting about abominations of nature and other nonsense. As though a living gallery was worse than cars and aeroplanes. At least my work would merge back into the earth in the end.

One of the guard's radios hissed out a message. Too quiet to hear, but the guards were running now.

A dull boom reverberated through the meat. Every door muscle spasmed.

I sprinted past the guards towards the sound. A stream of blood flowed from a torn door muscle, but it was minor compared with the damage in the room. It had been a precise attack. The bomber set it off by the wall nearest the control centre, taking out the meat and a chunk of the computer. The muscles would survive on their own for a while, but without a computer to coordinate repair, *Dead Meat II* would bleed to death.

The guards arrived behind me. "Ma'am?" one asked.



JACEY

"Get the back-up systems."

They left and I turned my attention to the wounds. Even without control, the nanobots would run a basic routine of sealing meat to meat. That was something I had. The bomber, and several guards, were splattered over the walls. I picked up a mangled arm and placed it against the bleeding door muscle, until the arm fused over the damage.

I started the same process on the large hole, but before long the blood was up to my ankles. Too much for the pores to take back in.

"It'll be an hour, ma'am," a guard reported from the door, with a glance at the arm lodged there.

When I was younger, in the days when I thought that painting was the ultimate expression, I'd have been angry. Now, anger was just another seam, keeping me from what I had to do. I climbed into the wound and rested my head against the damaged flesh. The guard's shouts faded as my wound healed around me.

My eyes opened in every room. I marvelled at the seamless perfection of the artwork wandering my corridors, in their shades of wonder and fear.

In my entrance room, I greeted my art. "Welcome to *Dead Meat III*." ■

Polenth Blake *(www.polenthblake.com) lives in England with her pet cockroaches, and prefers her meat cooked.*

nature

A key thread in your scientific career

Since 1869 we've made it our mission to share the greatest developments in science and advance the discussion among the global scientific community. Whether in print, online or mobile, *Nature* is always at the forefront of science and is your forum to read, watch, listen and engage with key research, news and opinion.

How will you access the next landmark discovery?

NATURE.COM/NATURE

nature publishing group npg

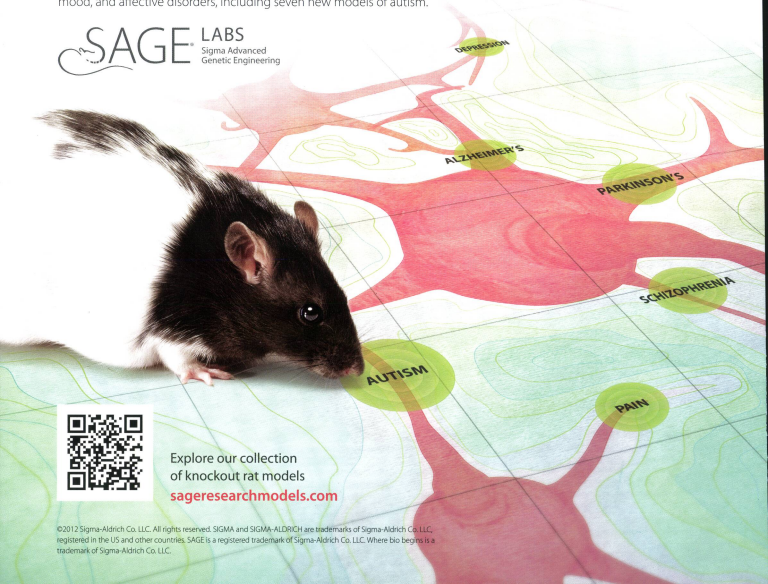
More translational and predictive animal models.

Plot your pathway for breakthroughs in neuroscience with knockout rat models from SAGE® Labs.

Biosynaptic.

Advancements in neuroscience are yours to discover with targeted knockout rats from SAGE Labs. Our comprehensive suite of rat models can help move your neuroscience research forward. Map your next breakthrough with smarter rat models for studying neurodegenerative, mood, and affective disorders, including seven new models of autism.

SAGE® LABS
Sigma Advanced Genetic Engineering



Explore our collection of knockout rat models
sageresearchmodels.com

©2012 Sigma-Aldrich Co. LLC. All rights reserved. SIGMA and SIGMA-ALDRICH are trademarks of Sigma-Aldrich Co. LLC, registered in the US and other countries. SAGE is a registered trademark of Sigma-Aldrich Co. LLC. Where bio begins is a trademark of Sigma-Aldrich Co. LLC.